PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K 38/00 | A2

(11) International Publication Number: WO 99/10004

(43) International Publication Date: 4 March 1999 (04.03.99)

(21) International Application Number: PCT/US98/17922

(22) International Filing Date: 28 August 1998 (28.08.98)

(30) Priority Data:
60/057,394 29 August 1997 (29.08.97) US

(71) Applicant: ONTOGENY, INC. [US/US]; 45 Moulton Street, Cambridge, MA 02138 (US).

(72) Inventors: BLADGEN, Chris, S.; 26–29 Drury Lane, London WC2B 5RL (GB). CURRIE, Peter, D.; 26–29 Drury Lane, London WC2B 5RL (GB). INGHAM, Philip, W.; Firth Court, Western Bank, Sheffield S10 2TN (GB). HUGHES, Simon, M.; 26–29 Drury Lane, London WC2B 5RL (GB).

(74) Agents: VINCENT, Matthew, P. et al.; Foley, Hoag & Eliot, LLP, One Post Office Square, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: REGULATION OF MUSCLE TISSUES BY HEDGEHOG-LIKE POLYPEPTIDES, AND FORMULATIONS AND USES RELATED THERETO

(57) Abstract

The present application relates to a method for modulating the formation and/or maintenance of muscle tissue by ectopically contacting muscle cells, especially muscle stem/progenitor cells, *in vitro* or *in vivo*, with a hedgehog therapeutic or ptc therapeutic in an amount effective to alter the growth state the treated cells, e.g., relative to the absence of administration of the hedgehog therapeutic or ptc therapeutic.

BC

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Regulation of Muscle Tissues by Hedgehog-like Polypeptides, and Formulations and Uses Related Thereto

BACKGROUND OF THE INVENTION

Since nuclei in muscle fibers of vertebrate animals are incapable of DNA synthesis or mitotic division, increases in muscle fiber numbers or in numbers of muscle fiber nuclei are due to proliferation and subsequent differentiation of skeletal muscle precursor cells known as "myoblasts." In adults, myoblasts remain as a mitotically quiescent reserve precursor population which can, upon muscle injury, re-enter the cell cycle, undergo several rounds of proliferation, and subsequently differentiate and permanently exit the from the cell cycle. Upon differentiation, differentiated myoblasts ("myocytes") acquire the ability to fuse with one another or with muscle fibers, and also commence coordinate expression of a large set of muscle-specific myofibrillar and contractile proteins (e.g., muscle myosins and actin, troponin, tropomyosin, etc.).

Muscle tissue can grow by several different mechanisms which are controlled by different trophic factors. Muscle tissue can grow by hypertrophy, an enlargement of or increase in mass or size of muscle fibers, or by hyperplasia, an increase in the numbers of fibers or in the numbers of muscle nuclei, or by a combination of these two processes. Growth factors that act on skeletal muscle tissue can be divided into two broad groups. The factors that stimulate proliferation of myoblasts usually inhibit differentiation of myoblasts and inhibit the expression and action of the muscle regulatory transcription factors (MRFs). Conversely, the factors that stimulate differentiation of myoblasts usually stimulate expression of the MRFs and can contribute to muscle hypertrophy. Most pharmacologic agents currently under consideration as muscle trophic factors act to stimulate muscle hypertrophy. Such hypertrophic factors include, for example, growth hormone (GH) or insulin-like growth factor-I (IGF-I). Muscle hypertrophy can be assessed by the measurement of muscle fiber diameter in vivo or in vitro, or by the measurement of the accretion of the muscle-specific myofibrillar and contractile proteins.

Clinically, a decline in such skeletal muscle tissue mass, or muscle atrophy, is an important contributor to frailty in older individuals. In human males, muscle mass declines by one-third between the ages of 50 and 80. In older adults, extended hospitalization can result in further disuse atrophy leading to a potential loss of the ability for independent living and to a cascade of physical decline. Moreover, the physical aging process profoundly affects body composition, including significant reductions in lean

body mass and increases in central adiposity. The changes in overall adiposity and fat distribution appear to be important factors in many common "age-related" disorders such as hypertension, glucose intolerance and diabetes, dyslipidemia, and atherosclerotic cardiovascular disease. In addition, it is possible that the age-associated decrement in muscle mass, and subsequently in strength and endurance, may be a critical determinant for functional loss, dependence and disability. Muscle weakness is also a major factor prediposing the elderly to falls and the resulting morbidity and mortality. Complications from falls constitute the sixth leading cause of death among people over 65 years of age.

It is a goal of the present invention to provide definition of how spatial information in the early somite generates muscle tissues.

SUMMARY OF THE INVENTION

One aspect of the present application relates to a method for regulating the formation and/or maintenance of muscle tissue by ectopically contacting muscle cells, especially muscle stem/progenitor cells, *in vitro* or *in vivo*, with a hedgehog therapeutic or ptc therapeutic in an amount effective to alter the growth state the treated cells, e.g., relative to the absence of administeration of the hedgehog therapeutic or ptc therapeutic.

Wherein the subject method is carried out using a *hedgehog* therapeutic, the *hedgehog* therapeutic preferably a polypeptide including a *hedgehog* portion comprising at least a bioactive extracellular portion of a *hedgehog* protein, e.g., the *hedgehog* portion includes at least 50, 100 or 150 (contiguous) amino acid residues of an N-terminal half of a *hedgehog* protein.In preferred embodiments, the *hedgehog* portion includes at least a portion of the *hedgehog* protein corresponding to a 19kd fragment of the extracellular domain of a *hedgehog* protein.

In preferred embodiments, the *hedgehog* portion has an amino acid sequence at least 60, 75, 85, or 95 percent identical with a hedgehog protein of any of SEQ ID Nos. 10-18 or 20, though sequences identical to those sequence listing entries are also contemplated as useful in the present method. The *hedgehog* portion can be encoded by a nucleic acid which hybridizes under stringent conditions to a nucleic acid sequence of any of SEQ ID Nos. 1-9 or 19, e.g., the *hedgehog* portion can be encoded by a vertebrate *hedgehog* gene, especially a human *hedgehog* gene.

In other embodiments, the subject method can be carried out by administering a gene activation construct, wherein the gene activation construct is deigned to recombine

with a genomic *hedgehog* gene of the patient to provide a heterologous transcriptional regulatory sequence operatively linked to a coding sequence of the *hedgehog* gene.

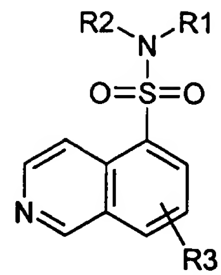
In still other embodiments, the subject method can be practiced with the administration of a gene therapy construct encoding a *hedgehog* polypeptide. For instance, the gene therapy construct can be provided in a composition selected from a group consisting of a recombinant viral particle, a liposome, and a poly-cationic nucleic acid binding agent,

In yet other embodiments, the subject method can be carried out using a ptc therapeutic. An exemplary ptc therapeutic is a small organic molecule which binds to a *patched* protein and derepresses *patched*-mediated inhibition of mitosis, e.g., a molecule which binds to *patched* and mimics *hedgehog*-mediated *patched* signal transduction, which binds to *patched* and regulates *patched*-dependent gene expression. For instance, the binding of the ptc therapeutic to *patched* may result in upregulation of patched and/or gli expression.

In a more generic sense, the *ptc* therapeutic can be a small organic molecule which interacts with muscle cells to induce *hedgehog*-mediated *patched* signal transduction, such as by altering the localization, protein-protein binding and/or enzymatic activity of an intracellular protein involved in a *patched* signal pathway. For instance, the *ptc* therapeutic may alter the level of expression of a *hedgehog* protein, a patched protein or a protein involved in the intracellular signal transduction pathway of *patched.*

In certain embodiments, the *ptc* therapeutic is an antisense construct which inhibits the expression of a protein which is involved in the signal transduction pathway of *patched* and the expression of which antagonizes *hedgehog*-mediated signals. The antisense construct is perferably an oligonucleotide of about 20-30 nucleotides in length and having a GC content of at least 50 percent.

In other embodiments, the *ptc* therapeutic is an inhibitor of protein kinase A (PKA), such as a 5-isoquinolinesulfonamide. The PKA inhibitor can be a cyclic AMP analog. Exemplary PKA inhibitors include N-[2-((p-bromocinnamyl)amino)ethyl]-5-isoquinolinesulfonamide, 1-(5-isoquinoline-sulfonyl)-2-methylpiperazine, KT5720, 8-bromo-cAMP, dibutyryl-cAMP and PKA Heat Stable Inhibitor isoform α. Another exemplary PKA inhibitor is represented in the general formula:



wherein,

R_1 and R_2 each can independently represent hydrogen, and as valence and stability permit a lower alkyl, a lower alkenyl, a lower alkynyl, a carbonyl (such as a carboxyl, an ester, a formate, or a ketone), a thiocarbonyl (such as a thioester, a thioacetate, or a thioformate), an amino, an acylamino, an amido, a cyano, a nitro, an azido, a sulfate, a sulfonate, a sulfonamido, $-(CH_2)_m-R_8$, $-(CH_2)_m-OH$, $-(CH_2)_m$-O-lower alkyl, $-(CH_2)_m$-O-lower alkenyl, $-(CH_2)_n-O-(CH_2)_m-R_8$, $-(CH_2)_m-SH$, $-(CH_2)_m$-S-lower alkyl, $-(CH_2)_m$-S-lower alkenyl, $-(CH_2)_n-S-(CH_2)_m-R_8$, or

R_1 and R_2 taken together with N form a heterocycle (substituted or unsubstituted);

R_3 is absent or represents one or more substitutions to the isoquinoline ring such as a lower alkyl, a lower alkenyl, a lower alkynyl, a carbonyl (such as a carboxyl, an ester, a formate, or a ketone), a thiocarbonyl (such as a thioester, a thioacetate, or a thioformate), an amino, an acylamino, an amido, a cyano, a nitro, an azido, a sulfate, a sulfonate, a sulfonamido, $-(CH_2)_m-R_8$, $-(CH_2)_m-OH$, $-(CH_2)_m$-O-lower alkyl, $-(CH_2)_m$-O-lower alkenyl, $-(CH_2)_n-O-(CH_2)_m-R_8$, $-(CH_2)_m-SH$, $-(CH_2)_m$-S-lower alkyl, $-(CH_2)_m$-S-lower alkenyl, $-(CH_2)_n-S-(CH_2)_m-R_8$;

R_8 represents a substituted or unsubstituted aryl, aralkyl, cycloalkyl, cycloalkenyl, or heterocycle; and

n and m are independently for each occurrence zero or an integer in the range of 1 to 6.

DETAILED DESCRIPTION OF THE INVENTION

All vertebrates have two classes of muscle fibers: slow and fast. Slow fibers have low-force long-duration contractions because they express myosin isoforms that are specialized for slow contraction and an oxidative metabolism. Fast fibers have distinct fast myosins and glycolytic metabolism, ideal for high-force short-duration contractions.

Each muscle has a specific mix and spatial array of fast and slow fibers from their formation early in development. How such arrays are patterned has previously been unknown, though evidence for two contrasting models existed. In one view, proliferative myoblasts are intrinsically committed to form either fast or slow fibers and accumulate in appropriate regions of the embryo. Clonal cell analysis in chick shows that myoblasts are heterogeneous prior to their differentiation: some are specialized to form slow muscle, whereas others form fast (DiMario et al., 1993; Miller and Stockdale, 1986; Schafer et al., 1987). Alternatively, naive myoblasts could be instructed by their environment to express specific isoforms of muscle proteins at the time of differentiation, as occurs in postnatal rodent muscles (Hughes and Blau, 1992). A resolution of this issue is suggested by studies in Drosophila where local extrinsic signals induce commitment of muscle founder myoblasts to the formation of a particular type of muscle in each location (Baylies et al., 1995). In the following examples, we examine vertebrate muscle patterning in the zebrafish somite. We show that the secreted glycoprotein Sonic hedgehog (SHH) regulates the decision between fast and slow muscle formation and we suggest this decision involves induction of a specifically slow myoblast type.

I. Overview

The present application is directed to the discovery that hedgehog gene products are involved in controlling the formation and/or maintenance of muscle tissue, especially slow (red) muscle. Certain aspects of the invention are directed to a preparations of hedgehog polypeptides, or other molecules which regulate *patched* or *smoothened* signalling, and their uses in stimulating muscle growth or differentiation in mammals. In particular embodiments, the invention is directed to the use of *hedgehog* polypeptides to stimulate muscle growth, differentiation or hypertrophy.

As described in the appended examples, *hedgehog* proteins are implicated in the proliferation and/or differentiation of myoblastic/myocytic cells and may provide early signals that regulate the differentiation of these or other precursor (stem) cells into muscle tissues. In general, the method of the present invention comprises contacting a muscle cells (collectively, muscle stem cells (myoblasts), and myocytic or other differentiated muscle cells), with an amount of a hedgehog therapeutic (defined infra) which produces a non-toxic response by the cell of (i) induction of of muscle tissue formation or maintenance of existing muscle tissue, or (ii) inhibition of muscle tissue formation, depending on the whether the hedgehog therapeutic is a sufficient hedgehog agonist or hedgehog antagonist. The subject method can be carried out on muscle cells

which may be either dispersed in culture or a part of an intact tissue or organ. Moreover, the method can be performed on cells which are provided in culture (*in vitro*), or on cells in a whole animal (*in vivo*).

In one aspect, the present invention provides pharmaceutical preparations and methods for controlling the formation of myoblastic-derived tissue utilizing, as an active ingredient, a *hedgehog* polypeptide or a mimetic thereof. The invention also relates to methods of controlling the functional performance of an muscle-derived tissue by use of the pharmaceutical preparations of the invention.

The *hedgehog* formulations of the present invention may be used as part of regimens in the treatment or prevention of disorders of, or surgical or cosmetic repair of, such muscle tissues. For instance, the subject method can be used for treating atrophy, or wasting, in particular, skeletal muscle atrophy and cardiac muscle atrophy. In addition, certain diseases wherein the muscle tissue is damaged, is abnormal or has atrophied, are treatable using the invention, such as, for example, normal aging, disuse atrophy, wasting or cachexia, and various secondary disorders associated with age and the loss of muscle mass, such as hypertension, glucose intolerance and diabetes, dyslipidemia and atherosclerotic cardiovascular disease. In addition, the therapeutic preparatios of the present invention may be used to treat rhabdomyosarcomas by regulating myoblast differentiation. The invention also is directed to the treatment of certain cardiac insufficiencies, such as congestive heart failure. The treatment of muscular myopathies such as muscular dystrophies is also embodied in the invention.

In certain embodiments, the subject compositions can be used to inhibit, rather than promote, growth of myoblastic-derived tissue. For instance, certain of the compositions disclosed herein may be applied to the treatment or prevention of a variety hyperplastic or neoplastic conditions affecting muscle tissue. The method can find application for the treatment or prophylaxis of, e.g., invasive muscle tumors and myoblastic sarcomas.

The subject *hedgehog* treatments are effective on both human and animal subjects afflicted with these conditions. Animal subjects to which the invention is applicable extend to both domestic animals and livestock, raised either as pets or for commercial purposes. Examples are dogs, cats, cattle, horses, sheep, hogs and goats.

Still another aspect of the present invention provides a method of stimulating the growth and regulating the differentiation of muscle cells and tissues in culture.

Without wishing to be bound by any particular theory, the induction of muscle formation by hedgehog proteins may be due at least in part to the ability of these proteins

to antagonize (directly or indirectly) *patched*-mediated regulation of gene expression and other physiological effects mediated by that protein. The *patched* gene product, a cell surface protein, is understood to signal through a pathway which causes transcriptional repression of members of the Wnt and Dpp/BMP families of morphogens, proteins which impart positional information. In development of the CNS and patterning of limbs in vertebrates, the introduction of *hedgehog* relieves (derepresses) this inhibition conferred by *patched*, allowing expression of particular gene programs.

Recently, it has been reported that mutations in the human version of *patched*, a gene first identified in a fruit fly developmental pathway, cause a hereditary skin cancer and may contribute to sporadic skin cancers. See, for example, Hahn et al. (1996) *Cell* 86:841-851; and Johnson et al. (1996) *Science* 272:1668-1671. The demonstraction that nevoid basal-cell carcinoma (NBCC) results from mutations in the human *patched* gene provided an example of the roles *patched* plays in post-embryonic deveolpment. These observations have led the art to understand one activity of *patched* to be a tumor suppressor gene, which may act by inhibiting proliferative signals from *hedgehog*. Our observations set forth below reveal potential new roles for the *hedgehog/patched* pathway in maintenance of muscle cell proliferation and differentiation. Accordingly, the present invention contemplates the use of other agents which are capable of mimicking the effect of the *hedgehog* protein on *patched* signalling, e.g., as may be identified from the drug screening assays described below.

II. Definitions

For convience, certain terms employed in the specfication, examples, and appended claims are collected here.

The term "hedgehog therapeutic" refers to various forms of hedgehog polypeptides, as well as peptidomimetics, which can modulate the proliferation/differentiation state of muscle cells by, as will be clear from the context of individual examples, mimicing or potentiating (agonizing) or inhibiting (antagonizing) the effects of a naturally-occurring *hedgehog* protein. A *hedgehog* therapeutic which mimics or potentiates the activity of a wild-type hedgehog protein is a "hedgehog agonist". Conversely, a *hedgehog* therapeutic which inhibits the activity of a wild-type hedgehog protein is a "hedgehog antagonist".

In particular, the term "hedgehog polypeptide" encompasses preparations of *hedgehog* proteins and peptidyl fragments thereof, both agonist and antagonist forms as the specific context will make clear.

As used herein the term "bioactive fragment of a *hedgehog* protein" refers to a fragment of a full-length *hedgehog* polypeptide, wherein the fragment specifically agonizes or antagonizes inductive events mediated by wild-type *hedgehog* proteins. The *hedgehog* biactive fragment preferably is a soluble extracellular portion of a *hedgehog* protein, where solubility is with reference to physiologically compatible solutions. Exemplary bioactive fragments are described in PCT publications WO 95/18856 and WO 96/17924.

The term "ptc therapeutic" refers to agents which either (i) mimic the effect of *hedgehog* proteins on *patched* signalling, e.g., which antagonize the cell-cycle inhibitory activity of *patched*, or (ii) activate or potentiate patched signalling. In other embodiments, the ptc therapeutic can be a *hedgehog* antagonist. The ptc therapeutic can be, e.g., a peptide, a nucleic acid, a carbohydrate, a small organic molecule, or natural product extract (or fraction thereof).

A "proliferative" form of a hedgehog or ptc therapeutic is one which induces proliferation of muscle cells, particularly muscle stem cells. Conversely, an "antiproliferative" form of a hedgehog or ptc therapeutic is one which inhibits proliferation of an muscle cells, preferably in a non-toxic manner, e.g., by promoting or maintaining a differentiated phenotype or otherwise promoting quiescence.

As used herein, "myoblast cultures" refers to cultures that contain cycling skeletal muscle precursors, and are considered distinct from "muscle fiber cultures" which are derived from myoblast cultures that are allowed to undergo differentiation and fusion to form multinucleated muscle fibers. The term "myogenic culture" is a generic term that refers to both kinds of cultures. The term "myocyte" refers to a differentiated, post-mitotic, muscle cell that has not yet undergone fusion, and thus represents, in general, a transient cell type under most conditions.

As used herein, "proliferating" and "proliferation" refer to cells undergoing mitosis.

As used herein, "transformed cells" refers to cells which have spontaneously converted to a state of unrestrained growth, i.e., they have acquired the ability to grow through an indefinite number of divisions in culture. Transformed cells may be characterized by such terms as neoplastic, anaplastic and/or hyperplastic, with respect to their loss of growth control.

As used herein, "immortalized cells" refers to cells which have been altered via chemical and/or recombinant means such that the cells have the ability to grow through an indefinite number of divisions in culture.

A "patient" or "subject" to be treated by the subject method can mean either a human or non-human animal.

An "effective amount" of, e.g., a hedgehog therapeutic, with respect to the subject method of treatment, refers to an amount of, e.g., a hedgehog polypeptide in a preparation which, when applied as part of a desired dosage regimen brings about a change in the rate of cell proliferation and/or the state of differentiation of a cell so as to produce an amount of muscle cell proliferation or differentiation according to clinically acceptable standards for the disorder to be treated or the cosmetic purpose.

The "growth state" of a cell refers to the rate of proliferation of the cell and the state of differentiation of the cell.

"Homology" and "identity" each refer to sequence similarity between two polypeptide sequences, with identity being a more strict comparison. Homology and identity can each be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same amino acid residue, then the polypeptides can be referred to as identical at that position; when the equivalent site is occupied by the same amino acid (e.g., identical) or a similar amino acid (e.g., similar in steric and/or electronic nature), then the molecules can be refered to as homologous at that position. A percentage of homology or identity between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40 percent identity, though preferably less than 25 percent identity, with an hedgeog sequence of the present invention.

The term "corresponds to", when referring to a particular polypeptide or nucleic acid sequence is meant to indicate that the sequence of interest is identical or homologous to the reference sequence to which it is said to correspond.

The terms "recombinant protein", "heterologous protein" and "exogenous protein" are used interchangeably throughout the specification and refer to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide is inserted into a suitable expression construct which is in turn used to transform a host cell to produce the heterologous protein. That is, the polypeptide is expressed from a heterologous nucleic acid.

A "chimeric protein" or "fusion protein" is a fusion of a first amino acid sequence encoding a *hedgehog* polypeptide with a second amino acid sequence defining a domain foreign to and not substantially homologous with any domain of *hh* protein. A chimeric protein may present a foreign domain which is found (albeit in a different protein) in an

organism which also expresses the first protein, or it may be an "interspecies", "intergenic", etc. fusion of protein structures expressed by different kinds of organisms. In general, a fusion protein can be represented by the general formula $(X)_n$-$(hh)_m$-$(Y)_n$, wherein *hh* represents all or a portion of the *hedgehog* protein, X and Y each independently represent an amino acid sequences which are not naturally found as a polypeptide chain contiguous with the *hedgehog* sequence, m is an integer greater than or equal to 1, and each occurrence of n is, independently, 0 or an integer greater than or equal to 1 (n and m are preferably no greater than 5 or 10).

III. Exemplary Applications of Method and Compositions

The subject method has wide applicability to the treatment or prophylaxis of disorders afflicting muscle tissue. In general, the method can be characterized as including a step of administering to an animal an amount of a *ptc* or hedgehog therapeutic effective to alter the proliferative state of a treated muscle tissue. The mode of administration and dosage regimens will vary depending on the muscle tissue(s) which is to be treated. Likewise, as described in further detail below, the use of a particular ptc or hedgehog therapeutic, e.g., an agonist or antagonist, will depend on whether proliferation of cells of the treated tissue is desired or intended to be prevented.

In one aspect, the invention is directed to a muscle-trophic factor, and its use in stimulating muscle growth or differentiation in mammals. Such stimulation of muscle growth is useful for treating atrophy, or wasting, in particular, skeletal muscle atrophy and cardiac muscle atrophy. In addition, certain diseases wherein the muscle tissue is damaged, is abnormal or has atrophied, are treatable using the invention, such as, for example, normal aging, disuse atrophy, wasting or cachexia, and various secondary disorders associated with age and the loss of muscle mass, such as hypertension, glucose intolerance and diabetes, dyslipidemia and atherosclerotic cardiovascular disease. The treatment of muscular myopathies such as muscular dystrophies is also embodied in the invention.

With denervation or disuse, skeletal muscles undergo rapid atrophy which leads to a profound decrease in size, protein content and contractile strength. This atrophy is an important component of many neuromuscular diseases in humans. In a clinical setting, compositions comprising the subject ptc and hedgehog therapeutics can be used for inhibiting muscle degeration, e.g., for decreasing the loss of muscle mass, such as part of a treatment for such muscle wasting disorders.

In preferred embodiments pharmaceutical compositions according to the invention are administered to patients suffering from a disorder, i.e., an abnormal physical condition, a disease or pathophysiological condition associated with abnormal and/or aberrant regulation of muscle tissue. The disorders for which the compositions of the invention are administered are preferably those which directly or indirectly produce a wasting (i.e., loss) of muscle mass, that is, a muscle wasting disorder. These include muscular dystrophies, cardiac cachexia, emphysema, leprosy, malnutrition, osteomalacia, child acute leukemia, AIDS cachexia and cancer cachexia.

The muscular dystrophies are genetic diseases which are characterized by progressive weakness and degeneration of muscle fibers without evidence of neural degeneration. In Duchenne muscular dystrophy (DMD) patients display an average of a 67% reduction in muscle mass, and in myotonic dystrophy, fractional muscle protein synthesis has been shown to be decreased by an average of 28%, without any corresponding decrease in non-muscle protein synthesis (possibly due to impaired end-organ response to anabolic hormones or substrates). Accelerated protein degradation has been demonstrated in the muscles of DMD patients. The subject method can be used as part of a therapeutic strategy for preventing, and in some instance reversing, the muscle wasting conditions associated with such dystrophies.

Severe congestive heart failure (CHF) is characterized by a "cardiac cachexia," i.e., a muscle protein wasting of both the cardiac and skeletal muscles, with an average 19% body weight decrease. The cardiac cachexia is caused by an increased rate of myofibrillar protein breakdown. The subject method can be used as part of a treatment for cardiac cachexia.

Emphysema is a chronic obstructive pulmonary disease, defined by an enlargement of the air spaces distal to the terminal non-respiratory bronchioles, accompanied by destructive changes of the alveolar walls. Clinical manifestations of reduced pulmonary functioning include coughing, wheezing, recurrent respiratory infections, edema, and functional impairment and shortened life-span. The efflux of tyrosine is increased by 47% in emphysematous patients. Also, whole body leucine flux remains normal, whole-body leucine oxidation is increased, and whole-body protein synthesis is decreased. The result is a decrease in muscle protein synthesis, accompanied by a decrease in whole body protein turnover and skeletal muscle mass. This decrease becomes increasingly evident with disease progression and long term deterioration. The subject ptc and hedgehog therapeutics may be used to prevent and/or reverse, the muscle wasting conditions associated with such diseases.

In diabetes mellitus, there is a generalized wasting of small muscle of the hands, which is due to chronic partial denervation (neuropathy). This is most evident and worsens with long term disease progression and severity. The subject method can be used as part of a therapeutic strategy for treatement of diabetes mellitus.

Leprosy is associated with a muscular wasting which occurs between the metacarpals of the thumb and index finger. Severe malnutrition is characterized by, inter alia, severe muscle wasting. The subject method can be used to treat muscle wasting effects of leprosy.

Osteomalacia is a nutritional disorder caused by a deficiency of vitamin D and calcium. It is referred to as "rickets" in children, and "osteomalacia" in adults. It is marked by a softening of the bones (due to impaired mineralization, with excess accumulation of osteoid), pain, tenderness, muscle wasting and weakness, anorexia, and overall weight loss. It can result from malnutrition, repeated pregnancies and lactation (exhausting or depleting vitamin D and calcium stores), and vitamin D resistance. The subject method can be used as part of a therapeutic strategy for treatment of osteomalacia.

In childhood acute leukemia there is protein energy malnutrition which results in skeletal muscle wasting. Studies have shown that some children exhibit the muscle wasting even before diagnosis of the leukemia, with an average 27% decrease in muscle mass. There is also a simultaneous 33%-37% increase in adipose tissue, resulting in no net change in relative body weight and limb circumference. Such patients may be amenable to treatment with a ptc or hedgehog therapeutic according to the method of the present invention.

Cancer cachexia is a complex syndrome which occurs with variable incidence in patients with solid tumors and hematological malignancies. Clinically, cancer cachexia is manifested as weight loss with massive depletion of both adipose tissue and lean muscle mass, and is one cause of death which results from cancer. Cancer cachexia patients have shorter survival times, and decreased response to chemotherapy. In addition to disorders which produce muscle wasting, other circumstances and conditions appear to be linked in some fashion with a decrease in muscle mass. Such afflictions include muscle wasting due to chronic back pain, advanced age, long term hospitalization due to illness or injury, alcoholism and corticosteroid therapy. The subject method can be used as part of a therapeutic strategy for preventing, and in some instance reversing, the muscle wasting conditions associated with such cancers.

Studies have shown that in severe cases of chronic lower back pain, there is paraspinal muscle wasting. Decreasing paraspinal muscle wasting alleviates pain and improves function. A course of treatment for disorder can include administration of a therapeutic amount of ptc or hedgehog therapeutics.

It is also believed that general weakness in old age is due to muscle wasting. As the body ages, an increasing proportion of skeletal muscle is replaced by fibrous tissue. The result is a significant reduction in muscle power, but only a marginal reduction in fat-free mass. The subject method can be used as part of a treatment and preventive strategies for preventing/reversing muscle wasting in elderly patients.

Studies have also shown that in patients suffering injuries or chronic illnesses, and hospitalized for long periods of time, there is long-lasting unilateral muscle wasting, with an average 31% decrease in muscle mass. Studies have also shown that this can be corrected with intensive physiotherapy. However, it may be more effective for many patients to at least augment such therapies with treatment by the subject method

In alcoholics there is wasting of the anterior tibial muscle. This proximal muscle damage is caused by neurogenic damage, namely, impaired glycolytic and phosphorylase enzyme activity. The damage becomes apparent and worsens the longer the duration of the alcohol abuse. Patients treated with corticosteroids experience loss of muscle mass. Such patients may also be amenable to treatment by the subject method.

The compounds of the invention can be used to alleviate the muscle mass loss resulting from the foregoing conditions, as well as others. Additionally, the ptc and hedgehog therapeutics of the present invention are useful in veterinary and animal husbandry applications to counter weight loss in animals, or to promote growth. For instance, the invention may also find use for increasing the efficiency of animal meat production. Specifically, animals may be fed or injected with a ptc or hedgehog therapeutic in order to increase overall skeletal muscle mass, e.g., to increase the weight of such farm animals as cows, pigs, sheep, chickens and salmon.

The maintenance of tissues and organs *ex vivo* is also highly desirable. Tissue replacement therapy is well established in the treatment of human disease. There are many situations where one may wish to transplant muscle cells, especially muscle stem cells, into a recipient host where the recipient's cells are missing, damaged or dysfunctional. muscle cells in muscle wasting disease. For example, transplantation of normal myoblasts may be useful to treat Duchenne muscular dystrophy and other muscle degeneration and wasting diseases. See, for example, Partridge (1991) *Muscle & Nerve*

14:197-212. In the case of myoblasts, they may be injected at various sites to treat muscle wasting diseases.

The subject method can be used to regulate the growth of muscle cells and tissue *in vitro*, as well as to accelerate the grafting of impanted muscle tissue to an animal host

In this regard, the present invention also concerns myoblast cultures which have been expanded by treatment with a hedgehog or other ptc therapeutic. In an illustrative embodiment, such a method comprises obtaining a muscle sample, preferably one including myoblasts; optionally treating the cell sample enzymically to separate the cells; culturing, in the presence of a hedgehog or ptc therapeutic.

IV. Exemplary hedgehog therapeutic compounds.

The *hedgehog* therapeutic compositions of the subject method can be generated by any of a variety of techniques, including purification of naturally occurring proteins, recombinantly produced proteins and synthetic chemistry. Polypeptide forms of the hedgehog therapeutics are preferably derived from vertebrate hedgehog proteins, e.g., have sequences corresponding to naturally occurring hedgehog proteins, or fragments thereof, from vertebrate organisms. However, it will be appreciated that the hedgehog polypeptide can correspond to a hedgehog protein (or fragment thereof) which occurs in any metazoan organism.

The various naturally-occurring *hedgehog* proteins from which the subject therapeutics can be derived are characterized by a signal peptide, a highly conserved N-terminal region, and a more divergent C-terminal domain. In addition to signal sequence cleavage in the secretory pathway (Lee, J.J. *et al.* (1992) *Cell* 71:33-50; Tabata, T. *et al.* (1992) *Genes Dev.* 2635-2645; Chang, D.E. *et al.* (1994) *Development* 120:3339-3353), *hedgehog* precursor proteins naturally undergo an internal autoproteolytic cleavage which depends on conserved sequences in the C-terminal portion (Lee *et al.* (1994) *Science* 266:1528-1537; Porter *et al.* (1995) *Nature* 374:363-366). This autocleavage leads to a 19 kD N-terminal peptide and a C-terminal peptide of 26-28 kD (Lee *et al.* (1992) *supra*; Tabata *et al.* (1992) *supra*; Chang *et al.* (1994) *supra;* Lee *et al.* (1994) *supra*; Bumcrot, D.A., *et al.* (1995) *Mol. Cell. Biol.* 15:2294-2303; Porter *et al.* (1995) *supra*; Ekker, S.C. *et al.* (1995) *Curr. Biol.* 5:944-955; Lai, C.J. *et al.* (1995) *Development* 121:2349-2360). The N-terminal peptide stays tightly associated with the surface of cells in which it was synthesized, while the C-terminal peptide is freely diffusible both *in vitro* and *in vivo* (Lee *et al.* (1994) *supra*; Bumcrot *et al.* (1995) *supra;* Mart', E. *et al.* (1995) *Development* 121:2537-2547; Roelink, H. *et al.* (1995) *Cell* 81:445-455). Cell surface

retention of the N-terminal peptide is dependent on autocleavage, as a truncated form of *hedgehog* encoded by an RNA which terminates precisely at the normal position of internal cleavage is diffusible *in vitro* (Porter *et al.* (1995) *supra*) and *in vivo* (Porter, J.A. *et al.* (1996) *Cell* 86, 21-34). Biochemical studies have shown that the autoproteolytic cleavage of the *hedgehog* precursor protein proceeds through an internal thioester intermediate which subsequently is cleaved in a nucleophilic substitution. It is suggested that the nucleophile is a small lipophilic molecule, more particularly cholesterol, which becomes covalently bound to the C-terminal end of the N-peptide (Porter *et al.* (1996) *supra*), tethering it to the cell surface.

The vertebrate family of *hedgehog* genes includes at least four members, e.g., paralogs of the single drosophila *hedgehog* gene (SEQ ID No. 19). Three of these members, herein referred to as Desert *hedgehog* (*Dhh*), Sonic *hedgehog* (*Shh*) and Indian *hedgehog* (*Ihh*), apparently exist in all vertebrates, including fish, birds, and mammals. A fourth member, herein referred to as tiggie-winkle *hedgehog* (*Thh*), appears specific to fish. According to the appended sequence listing, (see also Table 1) a chicken *Shh* polypeptide is encoded by SEQ ID No:1; a mouse *Dhh* polypeptide is encoded by SEQ ID No:2; a mouse *Ihh* polypeptide is encoded by SEQ ID No:3; a mouse *Shh* polypeptide is encoded by SEQ ID No:4 a zebrafish *Shh* polypeptide is encoded by SEQ ID No:5; a human *Shh* polypeptide is encoded by SEQ ID No:6; a human *Ihh* polypeptide is encoded by SEQ ID No:7; a human *Dhh* polypeptide is encoded by SEQ ID No. 8; and a zebrafish *Thh* is encoded by SEQ ID No. 9.

Table 1

Guide to *hedgehog* sequences in Sequence Listing

	Nucleotide	Amino Acid
Chicken *Shh*	SEQ ID No. 1	SEQ ID No. 10
Mouse *Dhh*	SEQ ID No. 2	SEQ ID No. 11
Mouse *Ihh*	SEQ ID No. 3	SEQ ID No. 12
Mouse *Shh*	SEQ ID No. 4	SEQ ID No. 13
Zebrafish *Shh*	SEQ ID No. 5	SEQ ID No. 14
Human *Shh*	SEQ ID No. 6	SEQ ID No. 15
Human *Ihh*	SEQ ID No. 7	SEQ ID No. 16
Human *Dhh*	SEQ ID No. 8	SEQ ID No. 17
Zebrafish *Thh*	SEQ ID No. 9	SEQ ID No. 18
Drosophila *HH*	SEQ ID No. 19	SEQ ID No. 20

In addition to the sequence variation between the various *hedgehog* homologs, the *hedgehog* proteins are apparently present naturally in a number of different forms, including a pro-form, a full-length mature form, and several processed fragments thereof.

The pro-form includes an N-terminal signal peptide for directed secretion of the extracellular domain, while the full-length mature form lacks this signal sequence.

As described above, further processing of the mature form occurs in some instances to yield biologically active fragments of the protein. For instance, *sonic hedgehog* undergoes additional proteolytic processing to yield two peptides of approximately 19 kDa and 27 kDa, the 19kDa fragment corresponding to an proteolytic N-terminal portion of the mature protein.

In addition to proteolytic fragmentation, the vertebrate *hedgehog* proteins can also be modified post-translationally, such as by glycosylation and/or addition of lipophilic moieties, such as stents, fatty acids, etc., though bacterially produced (e.g. unmodified) forms of the proteins still maintain certain of the bioactivities of the native protein. Bioactive fragments of *hedgehog* polypeptides of the present invention have been generated and are described in great detail in, e.g., PCT publications WO 95/18856 and WO 96/17924.

There are a wide range of lipophilic moieties with which hedgehog polypeptides can be derivatived. The term "lipophilic group", in the context of being attached to a hedgehog polypeptide, refers to a group having high hydrocarbon content thereby giving the group high affinity to lipid phases. A lipophilic group can be, for example, a relatively long chain alkyl or cycloalkyl (preferably n-alkyl) group having approximately 7 to 30 carbons. The alkyl group may terminate with a hydroxy or primary amine "tail". To further illustrate, lipophilic molecules include naturally-occurring and synthetic aromatic and non-aromatic moieties such as fatty acids, sterols, esters and alcohols, other lipid molecules, cage structures such as adamantane and buckminsterfullerenes, and aromatic hydrocarbons such as benzene, perylene, phenanthrene, anthracene, naphthalene, pyrene, chrysene, and naphthacene.

In one embodiment, the hedgehog polypeptide is modified with one or more sterol moieties, such as cholesterol. See, for example, PCT publication WO 96/17924. In certain embodiments, the cholesterol is preferably added to the C-terminal glycine were the hedgehog polypeptide corresponds to the naturally-occurring N-terminal proteolytic fragment.

In another embodiment, the hedgehog polypeptide can be modified with a fatty acid moiety, such as a myrostoyl, palmitoyl, stearoyl, or arachidoyl moiety. See, e.g., Pepinsky et al. (1998) J Biol. Chem 273: 14037.

In addition to those effects seen by cholesterol-addition to the C-terminus or fatty acid addition to the N-terminus of extracellular fragments of the protein, at least certain

of the biological activities of the hedgehog gene products are unexpectedly potentiated by derivativation of the protein with lipophilic moieties at other sites on the protein and/or by moieties other than cholesterol or fatty acids. Certain aspects of the invention are directed to the use of preparations of hedgehog polypeptides which are modified at sites other than N-terminal or C-terminal residues of the natural processed form of the protein, and/or which are modified at such terminal residues with lipophilic moieties other than a sterol at the C-terminus or fatty acid at the N-terminus.

Particularly useful as lipophilic molecules are alicyclic hydrocarbons, saturated and unsaturated fatty acids and other lipid and phospholipid moieties, waxes, cholesterol, isoprenoids, terpenes and polyalicyclic hydrocarbons including adamantane and buckminsterfullerenes, vitamins, polyethylene glycol or oligoethylene glycol, (C1-C18)-alkyl phosphate diesters, -O-CH2-CH(OH)-O-(C12-C18)-alkyl, and in particular conjugates with pyrene derivatives. The lipophilic moiety can be a lipophilic dye suitable for use in the invention include, but are not limited to, diphenylhexatriene, Nile Red, N-phenyl-1-naphthylamine, Prodan, Laurodan, Pyrene, Perylene, rhodamine, rhodamine B, tetramethylrhodamine, Texas Red, sulforhodamine, 1,1'-didodecyl-3,3,3',3'tetramethylindocarbocyanine perchlorate, octadecyl rhodamine B and the BODIPY dyes available from Molecular Probes Inc.

Other exemplary lipophilic moietites include aliphatic carbonyl radical groups include 1- or 2-adamantylacetyl, 3-methyladamant-1-ylacetyl, 3-methyl-3-bromo-1-adamantylacetyl, 1-decalinacetyl, camphoracetyl, camphaneacetyl, noradamantylacetyl, norbornaneacetyl, bicyclo[2.2.2.]-oct-5-eneacetyl, 1-methoxybicyclo[2.2.2.]-oct-5-ene-2-carbonyl, cis-5-norbornene-endo-2,3-dicarbonyl, 5-norbornen-2-ylacetyl, (1R)-(-)-myrtentaneacetyl, 2-norbornaneacetyl, anti-3-oxo-tricyclo[2.2.1.0<2,6>]-heptane-7-carbonyl, decanoyl, dodecanoyl, dodecenoyl, tetradecadienoyl, decynoyl or dodecynoyl.

The hedgehog polypeptide can be linked to the hydrophobic moiety in a number of ways including by chemical coupling means, or by genetic engineering.

There are a large number of chemical cross-linking agents that are known to those skilled in the art. For the present invention, the preferred cross-linking agents are heterobifunctional cross-linkers, which can be used to link the hedgehog polypeptide and hydrophobic moiety in a stepwise manner. Heterobifunctional cross-linkers provide the ability to design more specific coupling methods for conjugating to proteins, thereby reducing the occurrences of unwanted side reactions such as homo-protein polymers. A wide variety of heterobifunctional cross-linkers are known in the art. These include: succinimidyl 4-(N-maleimidomethyl) cyclohexane- 1-carboxylate (SMCC), m-

Maleimidobenzoyl-N- hydroxysuccinimide ester (MBS); N-succinimidyl (4-iodoacetyl) aminobenzoate (SIAB), succinimidyl 4-(p-maleimidophenyl) butyrate (SMPB), 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC); 4-succinimidyloxycarbonyl- a-methyl-a-(2-pyridyldithio)-tolune (SMPT), N-succinimidyl 3-(2-pyridyldithio) propionate (SPDP), succinimidyl 6-[3-(2-pyridyldithio) propionate] hexanoate (LC-SPDP). Those cross-linking agents having N-hydroxysuccinimide moieties can be obtained as the N-hydroxysulfosuccinimide analogs, which generally have greater water solubility. In addition, those cross-linking agents having disulfide bridges within the linking chain can be synthesized instead as the alkyl derivatives so as to reduce the amount of linker cleavage *in vivo*.

In addition to the heterobifunctional cross-linkers, there exists a number of other cross-linking agents including homobifunctional and photoreactive cross-linkers. Disuccinimidyl suberate (DSS), bismaleimidohexane (BMH) and dimethylpimelimidate·2 HCl (DMP) are examples of useful homobifunctional cross-linking agents, and bis-[ß-(4-azidosalicylamido)ethyl]disulfide (BASED) and N-succinimidyl-6(4'-azido-2'-nitrophenyl- amino)hexanoate (SANPAH) are examples of useful photoreactive cross-linkers for use in this invention. For a recent review of protein coupling techniques, see Means et al. (1990) *Bioconjugate Chemistry* 1:2-12, incorporated by reference herein.

One particularly useful class of heterobifunctional cross-linkers, included above, contain the primary amine reactive group, N-hydroxysuccinimide (NHS), or its water soluble analog N-hydroxysulfosuccinimide (sulfo-NHS). Primary amines (lysine epsilon groups) at alkaline pH's are unprotonated and react by nucleophilic attack on NHS or sulfo-NHS esters. This reaction results in the formation of an amide bond, and release of NHS or sulfo-NHS as a by-product.

Another reactive group useful as part of a heterobifunctional cross-linker is a thiol reactive group. Common thiol reactive groups include maleimides, halogens, and pyridyl disulfides. Maleimides react specifically with free sulfhydryls (cysteine residues) in minutes, under slightly acidic to neutral (pH 6.5-7.5) conditions. Halogens (iodoacetyl functions) react with -SH groups at physiological pH's. Both of these reactive groups result in the formation of stable thioether bonds.

The third component of the heterobifunctional cross-linker is the spacer arm or bridge. The bridge is the structure that connects the two reactive ends. The most apparent attribute of the bridge is its effect on steric hindrance. In some instances, a

longer bridge can more easily span the distance necessary to link two complex biomolecules. For instance, SMPB has a span of 14.5 angstroms.

Preparing protein-protein conjugates using heterobifunctional reagents is a two-step process involving the amine reaction and the sulfhydryl reaction. For the first step, the amine reaction, the protein chosen should contain a primary amine. This can be lysine epsilon amines or a primary alpha amine found at the N-terminus of most proteins. The protein should not contain free sulfhydryl groups. In cases where both proteins to be conjugated contain free sulfhydryl groups, one protein can be modified so that all sulfhydryls are blocked using for instance, N-ethylmaleimide (see Partis et al. (1983) J. Pro. Chem. 2:263, incorporated by reference herein). Ellman's Reagent can be used to calculate the quantity of sulfhydryls in a particular protein (see for example Ellman et al. (1958) Arch. Biochem. Biophys. 74:443 and Riddles et al. (1979) Anal. Biochem. 94:75, incorporated by reference herein).

The reaction buffer should be free of extraneous amines and sulfhydryls. The pH of the reaction buffer should be 7.0-7.5. This pH range prevents maleimide groups from reacting with amines, preserving the maleimide group for the second reaction with sulfhydryls.

The NHS-ester containing cross-linkers have limited water solubility. They should be dissolved in a minimal amount of organic solvent (DMF or DMSO) before introducing the cross-linker into the reaction mixture. The cross-linker/solvent forms an emulsion which will allow the reaction to occur.

The sulfo-NHS ester analogs are more water soluble, and can be added directly to the reaction buffer. Buffers of high ionic strength should be avoided, as they have a tendency to "salt out" the sulfo-NHS esters. To avoid loss of reactivity due to hydrolysis, the cross-linker is added to the reaction mixture immediately after dissolving the protein solution.

The reactions can be more efficient in concentrated protein solutions. The more alkaline the pH of the reaction mixture, the faster the rate of reaction. The rate of hydrolysis of the NHS and sulfo-NHS esters will also increase with increasing pH. Higher temperatures will increase the reaction rates for both hydrolysis and acylation.

Once the reaction is completed, the first protein is now activated, with a sulfhydryl reactive moiety. The activated protein may be isolated from the reaction mixture by simple gel filtration or dialysis. To carry out the second step of the cross-linking, the sulfhydryl reaction, the lipophilic group chosen for reaction with maleimides,

activated halogens, or pyridyl disulfides must contain a free sulfhydryl. Alternatively, a primary amine may be modified with to add a sulfhydryl

In all cases, the buffer should be degassed to prevent oxidation of sulfhydryl groups. EDTA may be added to chelate any oxidizing metals that may be present in the buffer. Buffers should be free of any sulfhydryl containing compounds.

Maleimides react specifically with -SH groups at slightly acidic to neutral pH ranges (6.5-7.5). A neutral pH is sufficient for reactions involving halogens and pyridyl disulfides. Under these conditions, maleimides generally react with -SH groups within a matter of minutes. Longer reaction times are required for halogens and pyridyl disulfides.

The first sulfhydryl reactive-protein prepared in the amine reaction step is mixed with the sulfhydryl-containing lipophilic group under the appropriate buffer conditions. The conjugates can be isolated from the reaction mixture by methods such as gel filtration or by dialysis.

Exemplary activated lipophilic moieties for conjugation include: N-(1-pyrene)maleimide; 2,5-dimethoxystilbene-4'-maleimide, eosin-5-maleimide; fluorescein-5-maleimide; N-(4-(6-dimethylamino- 2-benzofuranyl)phenyl)maleimide; benzophenone-4-maleimide; 4-dimethylaminophenylazophenyl- 4'-maleimide (DABMI), tetramethylrhodamine-5-maleimide, tetramethylrhodamine-6-maleimide, Rhodamine RedTM C2 maleimide, N-(5-aminopentyl)maleimide, trifluoroacetic acid salt, N-(2-aminoethyl)maleimide, trifluoroacetic acid salt, Oregon GreenTM 488 maleimide, N-(2-((2-(((4-azido- 2,3,5,6-tetrafluoro)benzoyl) amino)ethyl)dithio)ethyl)maleimide (TFPAM-SS1), 2-(1-(3-dimethylaminopropyl) -indol-3-yl)-3-(indol-3-yl) maleimide (bisindolylmaleimide; GF 109203X), BODIPY® FL N-(2-aminoethyl)maleimide, N-(7-dimethylamino- 4-methylcoumarin-3-yl)maleimide (DACM), AlexaTM 488 C5 maleimide, AlexaTM 594 C5 maleimide, sodium saltN-(1-pyrene)maleimide, 2,5-dimethoxystilbene-4'-maleimide, eosin-5-maleimide, fluorescein-5-maleimide, N-(4-(6-dimethylamino- 2-benzofuranyl)phenyl)maleimide, benzophenone-4-maleimide, 4-dimethylaminophenylazophenyl- 4'-maleimide, 1-(2-maleimidylethyl)-4-(5- (4-methoxyphenyl)oxazol-2- yl)pyridinium methanesulfonate, tetramethylrhodamine-5-maleimide, tetramethylrhodamine-6-maleimide, Rhodamine RedTM C2 maleimide, N-(5-aminopentyl)maleimide, N-(2-aminoethyl)maleimide, N-(2-((2-(((4-azido- 2,3,5,6-tetrafluoro)benzoyl) amino)ethyl)dithio)ethyl)maleimide, 2-(1-(3-dimethylaminopropyl) -indol-3-yl)-3-(indol-3-yl) maleimide, N-(7-dimethylamino- 4-methylcoumarin-3-yl)maleimide (DACM), 11H-Benzo[a]fluorene, Benzo[a]pyrene.

In one embodiment, the hedgehog polypeptide can be derivatived using pyrene maleimide, which can be purchased from Molecular Probes (Eugene, Oreg.), e.g., N-(1-pyrene)maleimide or 1-pyrenemethyl iodoacetate (PMIA ester).

For those embodiments wherein the hydophobic moiety is a polypeptide, the modified hedgehog polypeptide of this invention can be constructed as a fusion protein, containing the hedgehog polypeptide and the hydrophobic moiety as one contiguous polypeptide chain.

In certain embodiments, the lipophilic moiety is an amphipathic polypeptide, such as magainin, cecropin, attacin, melittin, gramicidin S, alpha-toxin of Staph. aureus, alamethicin or a synthetic amphipathic polypeptide. Fusogenic coat proteins from viral particles can also be a convenient source of amphipathic sequences for the subject hedgehog proteins

Moreover, mutagenesis can be used to create modified *hh* polypeptides, e.g., for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*). Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition. Modified *hedgehog* polypeptides can also include those with altered post-translational processing relative to a naturally occurring *hedgehog* protein, e.g., altered glycosylation, cholesterolization, prenylation and the like.

In one embodiment, the hedgehog therapeutic is a polypeptide encodable by a nucleotide sequence that hybridizes under stringent conditions to a hedgehog coding sequence represented in one or more of SEQ ID Nos:1-7. Appropriate stringency conditions which promote DNA hybridization, for example, 6.0 x sodium chloride/sodium citrate (SSC) at about 45°C, followed by a wash of 2.0 x SSC at 50°C, are known to those skilled in the art or can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. For example, the salt concentration in the wash step can be selected from a low stringency of about 2.0 x SSC at 50°C to a high stringency of about 0.2 x SSC at 50°C. In addition, the temperature in the wash step can be increased from low stringency conditions at room temperature, about 22°C, to high stringency conditions at about 65°C.

As described in the literature, genes for other hedgehog proteins, e.g., from other animals, can be obtained from mRNA or genomic DNA samples using techniques well known in the art. For example, a cDNA encoding a *hedgehog* protein can be obtained by isolating total mRNA from a cell, e.g. a mammalian cell, e.g. a human cell, including embryonic cells. Double stranded cDNAs can then be prepared from the total mRNA,

and subsequently inserted into a suitable plasmid or bacteriophage vector using any one of a number of known techniques. The gene encoding a *hedgehog* protein can also be cloned using established polymerase chain reaction techniques.

Preferred nucleic acids encode a *hedgehog* polypeptide comprising an amino acid sequence at least 60% homologous or identical, more preferably 70% homologous or identical, and most preferably 80% homologous or identical with an amino acid sequence selected from the group consisting of SEQ ID Nos:8-14. Nucleic acids which encode polypeptides at least about 90%, more preferably at least about 95%, and most preferably at least about 98-99% homology or identity with an amino acid sequence represented in one of SEQ ID Nos:8-14 are also within the scope of the invention.

In addition to native *hedgehog* proteins, h*edgehog* polypeptides preferred by the present invention are at least 60% homologous or identical, more preferably 70% homologous or identical and most preferably 80% homologous or identical with an amino acid sequence represented by any of SEQ ID Nos:8-14. Polypeptides which are at least 90%, more preferably at least 95%, and most preferably at least about 98-99% homologous or identical with a sequence selected from the group consisting of SEQ ID Nos:8-14 are also within the scope of the invention. The only prerequisite is that the *hedgehog* polypeptide is capable of modulating the growth of muscle cells.

The term "recombinant protein" refers to a polypeptide of the present invention which is produced by recombinant DNA techniques, wherein generally, DNA encoding a *hedgehog* polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. Moreover, the phrase "derived from", with respect to a recombinant *hedgehog* gene, is meant to include within the meaning of "recombinant protein" those proteins having an amino acid sequence of a native *hedgehog* protein, or an amino acid sequence similar thereto which is generated by mutations including substitutions and deletions (including truncation) of a naturally occurring form of the protein.

The method of the present invention can also be carried out using variant forms of the naturally occurring *hedgehog* polypeptides, e.g., mutational variants.

As is known in the art, hedgehog polypeptides can be produced by standard biological techniques or by chemical synthesis. For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding the subject polypeptides can be cultured under appropriate conditions to allow expression of the peptide to occur. The polypeptide *hedgehog* may be secreted and isolated from a mixture of cells and medium containing the recombinant *hedgehog* polypeptide. Alternatively,

the peptide may be retained cytoplasmically by removing the signal peptide sequence from the recombinant *hedgehog* gene and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. The recombinant *hedgehog* polypeptide can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such peptide. In a preferred embodiment, the recombinant *hedgehog* polypeptide is a fusion protein containing a domain which facilitates its purification, such as an *hedgehog*/GST fusion protein. The host cell may be any prokaryotic or eukaryotic cell.

Recombinant *hedgehog* genes can be produced by ligating nucleic acid encoding an *hedgehog* protein, or a portion thereof, into a vector suitable for expression in either prokaryotic cells, eukaryotic cells, or both. Expression vectors for production of recombinant forms of the subject *hedgehog* polypeptides include plasmids and other vectors. For instance, suitable vectors for the expression of a *hedgehog* polypeptide include plasmids of the types: pBR322-derived plasmids, pEMBL-derived plasmids, pEX-derived plasmids, pBTac-derived plasmids and pUC-derived plasmids for expression in prokaryotic cells, such as *E. coli*.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach *et al.* (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in *E. coli* due to the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. In an illustrative embodiment, an *hedgehog* polypeptide is produced recombinantly utilizing an expression vector generated by sub-cloning the coding sequence of one of the *hedgehog* genes represented in SEQ ID Nos:1-7.

The preferred mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pcDNAI/amp, pcDNAI/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dhfr, pTk2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified

with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (pHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989) Chapters 16 and 17.

In some instances, it may be desirable to express the recombinant *hedgehog* polypeptide by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the ß-gal containing pBlueBac III).

When it is desirable to express only a portion of an *hedgehog* protein, such as a form lacking a portion of the N-terminus, i.e. a truncation mutant which lacks the signal peptide, it may be necessary to add a start codon (ATG) to the oligonucleotide fragment containing the desired sequence to be expressed. It is well known in the art that a methionine at the N-terminal position can be enzymatically cleaved by the use of the enzyme methionine aminopeptidase (MAP). MAP has been cloned from *E. coli* (Ben-Bassat et al. (1987) *J. Bacteriol.* 169:751-757) and *Salmonella typhimurium* and its *in vitro* activity has been demonstrated on recombinant proteins (Miller et al. (1987) *PNAS 84*:2718-1722). Therefore, removal of an N-terminal methionine, if desired, can be achieved either *in vivo* by expressing *hedgehog*-derived polypeptides in a host which produces MAP (e.g., *E. coli* or CM89 or *S. cerevisiae*), or *in vitro* by use of purified MAP (e.g., procedure of Miller et al., *supra*).

Alternatively, the coding sequences for the polypeptide can be incorporated as a part of a fusion gene including a nucleotide sequence encoding a different polypeptide. It is widely appreciated that fusion proteins can also facilitate the expression of proteins, and accordingly, can be used in the expression of the *hedgehog* polypeptides of the present invention. For example, *hedgehog* polypeptides can be generated as glutathione-S-transferase (GST-fusion) proteins. Such GST-fusion proteins can enable easy purification of the *hedgehog* polypeptide, as for example by the use of glutathione-derivatized matrices (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. (N.Y.: John Wiley & Sons, 1991)). In another embodiment, a fusion gene coding for a purification leader sequence, such as a poly-(His)/enterokinase cleavage site

sequence, can be used to replace the signal sequence which naturally occurs at the N-terminus of the *hedgehog* protein (e.g.of the pro-form, in order to permit purification of the poly(His)-*hedgehog* protein by affinity chromatography using a Ni^{2+} metal resin. The purification leader sequence can then be subsequently removed by treatment with enterokinase (e.g., see Hochuli et al. (1987) *J. Chromatography* 411:177; and Janknecht et al. *PNAS* 88:8972).

Techniques for making fusion genes are known to those skilled in the art. Essentially, the joining of various DNA fragments coding for different polypeptide sequences is performed in accordance with conventional techniques, employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel et al. John Wiley & Sons: 1992).

Hedgehog polypeptides may also be chemically modified to create *hedgehog* derivatives by forming covalent or aggregate conjugates with other chemical moieties, such as glycosyl groups, cholesterol, isoprenoids, lipids, phosphate, acetyl groups and the like. Covalent derivatives of *hedgehog* proteins can be prepared by linking the chemical moieties to functional groups on amino acid sidechains of the protein or at the N-terminus or at the C-terminus of the polypeptide.

For instance, *hedgehog* proteins can be generated to include a moiety, other than sequence naturally associated with the protein, that binds a component of the extracellular matrix and enhances localization of the analog to cell surfaces. For example, sequences derived from the fibronectin "type-III repeat", such as a tetrapeptide sequence R-G-D-S (Pierschbacher et al. (1984) *Nature* 309:30-3; and Kornblihtt et al. (1985) *EMBO* 4:1755-9) can be added to the *hedgehog* polypeptide to support attachment of the chimeric molecule to a cell through binding ECM components (Ruoslahti et al. (1987) *Science* 238:491-497; Pierschbacheret al. (1987) *J. Biol. Chem.* 262:17294-8.; Hynes (1987) *Cell* 48:549-54; and Hynes (1992) *Cell* 69:11-25).

In a preferred embodiment, the *hedgehog* polypeptide is isolated from, or is otherwise substantially free of, other cellular proteins, especially other extracellular or

cell surface associated proteins which may normally be associated with the *hedgehog* polypeptide, unless provided in the form of fusion protein with the *hedgehog* polypeptide. The term "substantially free of other cellular or extracellular proteins" (also referred to herein as "contaminating proteins") or "substantially pure preparations" or "purified preparations" are defined as encompassing preparations of *hedgehog* polypeptides having less than 20% (by dry weight) contaminating protein, and preferably having less than 5% contaminating protein. By "purified", it is meant that the indicated molecule is present in the substantial absence of other biological macromolecules, such as other proteins. The term "purified" as used herein preferably means at least 80% by dry weight, more preferably in the range of 95-99% by weight, and most preferably at least 99.8% by weight, of biological macromolecules of the same type present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 5000, can be present). The term "pure" as used herein preferably has the same numerical limits as "purified" immediately above.

As described above for recombinant polypeptides, isolated *hedgehog* polypeptides can include all or a portion of the amino acid sequences represented in any of SEQ ID Nos:10-18 or 20, or a homologous sequence thereto. Preferred fragments of the subject *hedgehog* proteins correspond to the N-terminal and C-terminal proteolytic fragments of the mature protein. Bioactive fragments of *hedgehog* polypeptides are described in great detail in PCT publications WO 95/18856 and WO 96/17924.

With respect to bioctive fragments of *hedgehog* polypeptide, preferred *hedgehog* therapeutics include at least 50 (contiguous) amino acid residues of a *hedgehog* polypeptide, more preferably at least 100 (contiguous), and even more preferably at least 150 (contiguous) residues.

Another preferred *hedgehog* polypeptide which can be included in the *hedgehog* therapeutic is an N-terminal fragment of the mature protein having a molecular weight of approximately 19 kDa.

Preferred human *hedgehog* proteins include N-terminal fragments corresponding approximately to residues 24-197 of SEQ ID No. 15, 28-202 of SEQ ID No. 16, and 23-198 of SEQ ID No. 17. By "corresponding approximately" it is meant that the sequence of interest is at most 20 amino acid residues different in length to the reference sequence, though more preferably at most 5, 10 or 15 amino acid different in length.

As described above for recombinant polypeptides, isolated *hedgehog* polypeptides can include all or a portion of the amino acid sequences represented in SEQ ID No:8, SEQ ID No:9, SEQ ID No:10, SEQ ID No:11, SEQ ID No:12, SEQ ID No:13

or SEQ ID No:14, or a homologous sequence thereto. Preferred fragments of the subject *hedgehog* proteins correspond to the N-terminal and C-terminal proteolytic fragments of the mature protein. Bioactive fragments of hedgehog polypeptides are described in great detail in PCT publications WO 95/18856 and WO 96/17924.

Still other preferred *hedgehog* polypeptides includes an amino acid sequence represented by the formula A-B wherein: (i) A represents all or the portion of the amino acid sequence designated by residues 1-168 of SEQ ID No:21; and B represents at least one amino acid residue of the amino acid sequence designated by residues 169-221 of SEQ ID No:21; (ii) A represents all or the portion of the amino acid sequence designated by residues 24-193 of SEQ ID No:15; and B represents at least one amino acid residue of the amino acid sequence designated by residues 194-250 of SEQ ID No:15; (iii) A represents all or the portion of the amino acid sequence designated by residues 25-193 of SEQ ID No:13; and B represents at least one amino acid residue of the amino acid sequence designated by residues 194-250 of SEQ ID No:13; (iv) A represents all or the portion of the amino acid sequence designated by residues 23-193 of SEQ ID No:11; and B represents at least one amino acid residue of the amino acid sequence designated by residues 194-250 of SEQ ID No:11; (v) A represents all or the portion of the amino acid sequence designated by residues 28-197 of SEQ ID No:12; and B represents at least one amino acid residue of the amino acid sequence designated by residues 198-250 of SEQ ID No:12; (vi) A represents all or the portion of the amino acid sequence designated by residues 29-197 of SEQ ID No:16; and B represents at least onc amino acid residue of the amino acid sequence designated by residucs 198-250 of SEQ ID No:16; or (vii) A represents all or the portion of the amino acid scquence designated by residues 23-193 of SEQ ID No. 17, and B represents at least one amino acid residue of the amino acid sequence designated by residues 194-250 of SEQ ID No. 17. In certain preferred embodiments, A and B together represent a contiguous polypeptide sequence designated sequence, A represents at least 25, 50, 75, 100, 125 or 150 (contiguous) amino acids of the designated sequence, and B represents at least 5, 10, or 20 (contiguous) amino acid residues of the amino acid sequence designated by corresponding entry in the sequence listing, and A and B together preferably represent a contiguous sequence corresponding to the sequence listing entry. Similar fragments from other *hedgehog* also contemplated, e.g., fragments which correspond to the preferred fragments from the sequence listing entries which are enumerated above. In preferred embodiments, the *hedgehog* polypeptide includes a C-terminal glycine (or other appropriate residue) which is derivatized with a cholesterol.

Isolated peptidyl portions of *hedgehog* proteins can be obtained by screening peptides recombinantly produced from the corresponding fragment of the nucleic acid encoding such peptides. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, a *hedgehog* polypeptide of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or preferably divided into overlapping fragments of a desired length. The fragments can be produced (recombinantly or by chemical synthesis) and tested to identify those peptidyl fragments which can function as either agonists or antagonists of a wild-type (e.g., "authentic") *hedgehog* protein. For example, Román et al. (1994) *Eur J Biochem* 222:65-73 describe the use of competitive-binding assays using short, overlapping synthetic peptides from larger proteins to identify binding domains.

The recombinant *hedgehog* polypeptides of the present invention also include homologs of the authentic *hedgehog* proteins, such as versions of those protein which are resistant to proteolytic cleavage, as for example, due to mutations which alter potential cleavage sequences or which inactivate an enzymatic activity associated with the protein. *Hedgehog* homologs of the present invention also include proteins which have been post-translationally modified in a manner different than the authentic protein. Exemplary derivatives of *hedgehog* proteins include polypeptides which lack N-glycosylation sites (e.g. to produce an unglycosylated protein), which lack sites for cholesterolization, and/or which lack N-terminal and/or C-terminal sequences.

Modification of the structure of the subject *hedgehog* polypeptides can also be for such purposes as enhancing therapeutic or prophylactic efficacy, or stability (e.g., *ex vivo* shelf life and resistance to proteolytic degradation *in vivo*). Such modified peptides, when designed to retain at least one activity of the naturally-occurring form of the protein, are considered functional equivalents of the *hedgehog* polypeptides described in more detail herein. Such modified peptides can be produced, for instance, by amino acid substitution, deletion, or addition.

It is well known in the art that one could reasonably expect that certain isolated replacements of amino acids, e.g., replacement of an amino acid residue with another related amino acid (i.e. isosteric and/or isoelectric mutations), can be carried out without major effect on the biological activity of the resulting molecule. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are can be divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) nonpolar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and

(4) uncharged polar = glycine, asparagine, glutamine, cysteine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. In similar fashion, the amino acid repertoire can be grouped as (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine histidine, (3) aliphatic = glycine, alanine, valine, leucine, isoleucine, serine, threonine, with serine and threonine optionally be grouped separately as aliphatic-hydroxyl; (4) aromatic = phenylalanine, tyrosine, tryptophan; (5) amide = asparagine, glutamine; and (6) sulfur -containing = cysteine and methionine. (see, for example, *Biochemistry*, 2nd ed., Ed. by L. Stryer, WH Freeman and Co.: 1981). Whether a change in the amino acid sequence of a peptide results in a functional *hedgehog* homolog (e.g. functional in the sense that it acts to mimic or antagonize the wild-type form) can be readily determined by assessing the ability of the variant peptide to produce a response in cells in a fashion similar to the wild-type protein, or competitively inhibit such a response. Polypeptides in which more than one replacement has taken place can readily be tested in the same manner.

It is specifically contemplated that the methods of the present invention can be carried using homologs of naturally occurring hedgehog proteins. In one embodiment, the invention contemplates using hedgehog polypeptides generated by combinatorial mutagenesis. Such methods, as are known in the art, are convenient for generating both point and truncation mutants, and can be especially useful for identifying potential variant sequences (e.g. homologs) that are functional in binding to a receptor for *hedgehog* proteins. The purpose of screening such combinatorial libraries is to generate, for example, novel *hedgehog* homologs which can act as either agonists or antagonist. To illustrate, *hedgehog* homologs can be engineered by the present method to provide more efficient binding to a cognate receptor, such as *patched*, yet still retain at least a portion of an activity associated with *hedgehog*. Thus, combinatorially-derived homologs can be generated to have an increased potency relative to a naturally occurring form of the protein. Likewise, *hedgehog* homologs can be generated by the present combinatorial approach to act as antagonists, in that they are able to mimic, for example, binding to other extracellular matrix components (such as receptors), yet not induce any biological response, thereby inhibiting the action of authentic *hedgehog* or *hedgehog* agonists. Moreover, manipulation of certain domains of *hedgehog* by the present method can provide domains more suitable for use in fusion proteins, such as one that incorporates portions of other proteins which are derived from the extracellular matrix and/or which bind extracellular matrix components.

To further illustrate the state of the art of combinatorial mutagenesis, it is noted that the review article of Gallop et al. (1994) *J Med Chem* 37:1233 describes the general

state of the art of combinatorial libraries as of the earlier 1990's. In particular, Gallop et al state at page 1239 "[s]creening the analog libraries aids in determining the minimum size of the active sequence and in identifying those residues critical for binding and intolerant of substitution". In addition, the Ladner et al. PCT publication WO90/02809, the Goeddel et al. U.S. Patent 5,223,408, and the Markland et al. PCT publication WO92/15679 illustrate specific techniques which one skilled in the art could utilize to generate libraries of *hedgehog* variants which can be rapidly screened to identify variants/fragments which retained a particular activity of the *hedgehog* polypeptides. These techniques are exemplary of the art and demonstrate that large libraries of related variants/truncants can be generated and assayed to isolate particular variants without undue experimentation. Gustin et al. (1993) *Virology* 193:653, and Bass et al. (1990) *Proteins: Structure, Function and Genetics* 8:309-314 also describe other exemplary techniques from the art which can be adapted as means for generating mutagenic variants of *hedgehog* polypeptides.

Indeed, it is plain from the combinatorial mutagenesis art that large scale mutagenesis of hedgehog proteins, without any preconceived ideas of which residues were critical to the biological function, and generate wide arrays of variants having equivalent biological activity. Indeed, it is the ability of combinatorial techniques to screen billions of different variants by high throughout analysis that removes any requirement of *a priori* understanding or knowledge of critical residues.

To illsutrate, the amino acid sequences for a population of *hedgehog* homologs or other related proteins are aligned, preferably to promote the highest homology possible. Such a population of variants can include, for example, *hedgehog* homologs from one or more species. Amino acids which appear at each position of the aligned sequences are selected to create a degenerate set of combinatorial sequences. In a preferred embodiment, the variegated library of *hedgehog* variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For instance, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential *hedgehog* sequences are expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display) containing the set of *hedgehog* sequences therein.

As illustrated in PCT publication WO 95/18856, to analyze the sequences of a population of variants, the amino acid sequences of interest can be aligned relative to sequence homology. The presence or absence of amino acids from an aligned sequence of a particular variant is relative to a chosen consensus length of a reference sequence, which can be real or artificial.

In an illustrative embodiment, alignment of exons 1, 2 and a portion of exon 3 encoded sequences (e.g. the N-terminal approximately 221 residues of the mature protein) of each of the *Shh* clones produces a degenerate set of *Shh* polypeptides represented by the general formula:

C-G-P-G-R-G-X(1)-G-X(2)-R-R-H-P-K-K-L-T-P-L-A-Y-K-Q-F-I-P-N-
V-A-E-K-T-L-G-A-S-G-R-Y-E-G-K-I-X(3)-R-N-S-E-R-F-K-E-L-T-P-N-
Y-N-P-D-I-I-F-K-D-E-E-N-T-G-A-D-R-L-M-T-Q-R-C-K-D-K-L-N-
X(4)-L-A-I-S-V-M-N-X(5)-W-P-G-V-X(6)-L-R-V-T-E-G-W-D-E-D-G-
H-H-X(7)-E-E-S-L-H-Y-E-G-R-A-V-D-I-T-T-S-D-R-D-X(8)-S-K-Y-G-
X(9)-L-X(10)-R-L-A-V-E-A-G-F-D-W-V-Y-Y-E-S-K-A-H-I-H-C-S-V-
K-A-E-N-S-V-A-A-K-S-G-G-C-F-P-G-S-A-X(11)-V-X(12)-L-X(13)-
X(14)-G-G-X(15)-K-X-(16)-V-K-D-L-X(17)-P-G-D-X(18)-V-L-A-A-D-
X(19)-X(20)-G-X(21)-L-X(22)-X(23)-S-D-F-X(24)-X(25)-F-X(26)-D-R
(SEQ ID No: 21

wherein each of the degenerate positions "X" can be an amino acid which occurs in that position in one of the human, mouse, chicken or zebrafish *Shh* clones, or, to expand the library, each X can also be selected from amongst amino acid residue which would be conservative substitutions for the amino acids which appear naturally in each of those positions. For instance, Xaa(1) represents Gly, Ala, Val, Leu, Ile, Phe, Tyr or Trp ; Xaa(2) represents Arg, His or Lys; Xaa(3) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(4) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(5) represents Lys, Arg, His, Asn or Gln; Xaa(6) represents Lys, Arg or His; Xaa(7) represents Ser, Thr, Tyr, Trp or Phe; Xaa(8) represents Lys, Arg or His; Xaa(9) represents Met, Cys, Ser or Thr; Xaa(10) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(11) represents Leu, Val, Met, Thr or Ser; Xaa(12) represents His, Phe, Tyr, Ser, Thr, Met or Cys; Xaa(13) represents Gln, Asn, Glu, or Asp; Xaa(14) represents His, Phe, Tyr, Thr, Gln, Asn, Glu or Asp; Xaa(15) represents Gln, Asn, Glu, Asp, Thr, Ser, Met or Cys; Xaa(16) represents Ala, Gly, Cys, Leu, Val or Met; Xaa(17) represents Arg, Lys, Met, Ile, Asn, Asp, Glu, Gln, Ser, Thr or Cys; Xaa(18) represents Arg, Lys, Met or Ile; Xaa(19) represents Ala, Gly, Cys, Asp, Glu, Gln, Asn, Ser, Thr or Met; Xaa(20) represents Ala, Gly, Cys, Asp, Asn, Glu or Gln; Xaa(21) represents Arg, Lys, Met, Ile, Asn, Asp, Glu or Gln; Xaa(22) represent Leu, Val, Met or Ile; Xaa(23) represents Phe, Tyr, Thr, His or Trp; Xaa(24) represents Ile, Val, Leu or Met; .Xaa(25) represents Met, Cys, Ile, Leu, Val, Thr or Ser; Xaa(26) represents Leu, Val, Met, Thr or Ser. In an even more expansive library, each X can be selected from any amino acid.

In similar fashion, alignment of each of the human, mouse, chicken and zebrafish *hedgehog* clones, can provide a degenerate polypeptide sequence represented by the general formula:

C-G-P-G-R-G-X(1)-X(2)-X(3)-R-R-X(4)-X(5)-X(6)-P-K-X(7)-L-X(8)-P-L-X(9)-Y-K-Q-F-X(10)-P-X(11)-X(12)-X(13)-E-X(14)-T-L-G-A-S-G-X(15)-X(16)-E-G-X(17)-X(18)-X(19)-R-X(20)-S-E-R-F-X(21)-X(22)-L-T-P-N-Y-N-P-D-I-I-F-K-D-E-E-N-X(23)-G-A-D-R-L-M-T-X(24)-R-C-K-X(25)-X(26)-X(27)-N-X(28)-L-A-I-S-V-M-N-X(29)-W-P-G-V-X(30)-L-R-V-T-E-G-X(31)-D-E-D-G-H-H-X(32)-X(33)-X(34)-S-L-H-Y-E-G-R-A-X(35)-D-I-T-T-S-D-R-D-X(36)-X(37)-K-Y-G-X(38)-L-X(39)-R-L-A-V-E-A-G-F-D-W-V-Y-Y-E-S-X(40)-X(41)-H-X(42)-H-X(43)-S-V-K-X(44)-X(45) (SEQ IDNo:22

wherein, as above, each of the degenerate positions "X" can be an amino acid which occurs in a corresponding position in one of the wild-type clones, and may also include amino acid residue which would be conservative substitutions, or each X can be any amino acid residue. In an exemplary embodiment, Xaa(1) represents Gly, Ala, Val, Leu, Ile, Pro, Phe or Tyr; Xaa(2) represents Gly, Ala, Val, Leu or Ile; Xaa(3) represents Gly, Ala, Val, Leu, Ile, Lys, His or Arg; Xaa(4) represents Lys, Arg or His; Xaa(5) represents Phe, Trp, Tyr or an amino acid gap; Xaa(6) represents Gly, Ala, Val, Leu, Ile or an amino acid gap; Xaa(7) represents Asn, Gln, His, Arg or Lys; Xaa(8) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(9) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(10) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(11) represents Ser, Thr, Gln or Asn; Xaa(12) represents Met, Cys, Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(13) represents Gly, Ala, Val, Leu, Ile or Pro; Xaa(14) represents Arg, His or Lys; Xaa(15) represents Gly, Ala, Val, Leu, Ile, Pro, Arg, His or Lys; Xaa(16) represents Gly, Ala, Val, Leu, Ile, Phe or Tyr; Xaa(17) represents Arg, His or Lys; Xaa(18) represents Gly, Ala, Val, Leu, Ile, Ser or Thr; Xaa(19) represents Thr or Ser; Xaa(20) represents Gly, Ala, Val, Leu, Ile, Asn or Gln; Xaa(21) represents Arg, His or Lys; Xaa(22) represents Asp or Glu; Xaa(23) represents Ser or Thr; Xaa(24) represents Glu, Asp, Gln or Asn; Xaa(25) represents Glu or Asp; Xaa(26) represents Arg, His or Lys; Xaa(27) represents Gly, Ala, Val, Leu or Ile; Xaa(28) represents Gly, Ala, Val, Leu, Ile, Thr or Ser; Xaa(29) represents Met, Cys, Gln, Asn, Arg, Lys or His; Xaa(30) represents Arg, His or Lys; Xaa(31) represents Trp, Phe, Tyr, Arg, His or Lys; Xaa(32) represents Gly, Ala, Val, Leu, Ile, Ser, Thr, Tyr or Phe; Xaa(33) represents Gln, Asn, Asp or Glu; Xaa(34) represents Asp or Glu; Xaa(35) represents Gly, Ala, Val, Leu, or Ile; Xaa(36) represents Arg, His or Lys; Xaa(37) represents Asn, Gln, Thr or Ser; Xaa(38) represents Gly, Ala, Val, Leu, Ile, Ser, Thr, Met or Cys; Xaa(39) represents Gly, Ala, Val, Leu, Ile, Thr or Ser; Xaa(40) represents Arg, His or Lys; Xaa(41) represents Asn, Gln, Gly, Ala, Val, Leu or Ile; Xaa(42) represents Gly, Ala, Val, Leu or Ile; Xaa(43) represents Gly, Ala, Val, Leu, Ile, Ser, Thr or Cys; Xaa(44) represents Gly, Ala, Val, Leu, Ile, Thr or Ser; and Xaa(45) represents Asp or Glu.

There are many ways by which the library of potential *hedgehog* homologs can be generated from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The purpose of a degenerate set of genes is to provide, in one mixture, all of the sequences encoding the desired set of potential *hedgehog* sequences. The synthesis of degenerate oligonucleotides is well known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

A wide range of techniques are known in the art for screening gene products of combinatorial libraries made by point mutations, and for screening cDNA libraries for gene products having a certain property. Such techniques will be generally adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of *hedgehog* homologs. The most widely used techniques for screening large gene libraries typically comprises cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the illustrative assays described below are amenable to high through-put analysis as necessary to screen large numbers of degenerate *hedgehog* sequences created by combinatorial mutagenesis techniques.

In one embodiment, the combinatorial library is designed to be secreted (e.g. the polypeptides of the library all include a signal sequence but no transmembrane or cytoplasmic domains), and is used to transfect a eukaryotic cell that can be co-cultured with muscle stem cells. A functional *hedgehog* protein secreted by the cells expressing the combinatorial library will diffuse to neighboring muscle cells and induce a particular biological response, such as proliferation or differentiation. The pattern of detection of such a change in phenotype will resemble a gradient function, and will allow the isolation (generally after several repetitive rounds of selection) of cells producing *hedgehog* homologs active as muscle-trophic agents. Likewise, *hedgehog* antagonists can be

selected in similar fashion by the ability of the cell producing a functional antagonist to protect neighboring cells (e.g., to inhibit proliferation) from the effect of wild-type *hedgehog* added to the culture media.

To illustrate, target muscle cells are cultured in 24-well microtitre plates. Other eukaryotic cells are transfected with the combinatorial *hedgehog* gene library and cultured in cell culture inserts (e.g. Collaborative Biomedical Products, Catalog #40446) that are able to fit into the wells of the microtitre plate. The cell culture inserts are placed in the wells such that recombinant *hedgehog* homologs secreted by the cells in the insert can diffuse through the porous bottom of the insert and contact the target cells in the microtitre plate wells. After a period of time sufficient for functional forms of a *hedgehog* protein to produce a measurable response in the target cells, such as growth state, the inserts are removed and the effect of the variant *hedgehog* proteins on the target cells determined. Cells from the inserts corresponding to wells which score positive for activity can be split and re-cultured on several inserts, the process being repeated until the active clones are identified.

In yet another screening assay, the candidate *hedgehog* gene products are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to associate with a *hedgehog*-binding moiety (such as the *patched* protein or other *hedgehog* receptor) via this gene product is detected in a "panning assay". Such panning steps can be carried out on cells cultured from embryos. For instance, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, fluorescently labeled molecules which bind *hedgehog* can be used to score for potentially functional *hedgehog* homologs. Cells can be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, separated by a fluorescence-activated cell sorter.

In an alternate embodiment, the gene library is expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at very high concentrations, large number of phage can be screened at one time. Second, since each infectious phage displays the combinatorial gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by another round of infection. The group of almost identical *E.coli* filamentous phages M13, fd, and fl are most often used in phage display libraries, as

either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) J. Biol. Chem. 267:16007-16010; Griffths et al. (1993) EMBO J 12:725-734; Clackson et al. (1991) Nature 352:624-628; and Barbas et al. (1992) PNAS 89:4457-4461).

In an illustrative embodiment, the recombinant phage antibody system (RPAS, Pharamacia Catalog number 27-9400-01) can be easily modified for use in expressing and screening *hedgehog* combinatorial libraries. For instance, the pCANTAB 5 phagemid of the RPAS kit contains the gene which encodes the phage gIII coat protein. The *hedgehog* combinatorial gene library can be cloned into the phagemid adjacent to the gIII signal sequence such that it will be expressed as a gIII fusion protein. After ligation, the phagemid is used to transform competent *E. coli* TG1 cells. Transformed cells are subsequently infected with M13KO7 helper phage to rescue the phagemid and its candidate *hedgehog* gene insert. The resulting recombinant phage contain phagemid DNA encoding a specific candidate *hedgehog*, and display one or more copies of the corresponding fusion coat protein. The phage-displayed candidate *hedgehog* proteins which are capable of binding an *hedgehog* receptor are selected or enriched by panning. For instance, the phage library can be applied to cells which express the *patched* protein and unbound phage washed away from the cells. The bound phage is then isolated, and if the recombinant phage express at least one copy of the wild type gIII coat protein, they will retain their ability to infect *E. coli*. Thus, successive rounds of reinfection of *E. coli*, and panning will greatly enrich for *hedgehog* homologs, which can then be screened for further biological activities in order to differentiate agonists and antagonists.

Combinatorial mutagenesis has a potential to generate very large libraries of mutant proteins, e.g., in the order of 10^{26} molecules. Combinatorial libraries of this size may be technically challenging to screen even with high throughput screening assays such as phage display. To overcome this problem, a new technique has been developed recently, recursive ensemble mutagenesis (REM), which allows one to avoid the very high proportion of non-functional proteins in a random library and simply enhances the frequency of functional proteins, thus decreasing the complexity required to achieve a useful sampling of sequence space. REM is an algorithm which enhances the frequency of functional mutants in a library when an appropriate selection or screening method is employed (Arkin and Yourvan, 1992, *PNAS USA* 89:7811-7815; Yourvan et al., 1992, *Parallel Problem Solving from Nature*, 2., In Maenner and Manderick, eds., Elsevir

Publishing Co., Amsterdam, pp. 401-410; Delgrave et al., 1993, *Protein Engineering* 6(3):327-331).

The invention also provides for reduction of the *hedgehog* protein to generate mimetics, e.g. peptide or non-peptide agents, which are able to disrupt binding of a *hedgehog* polypeptide of the present invention with an *hedgehog* receptor. Thus, such mutagenic techniques as described above are also useful to map the determinants of the *hedgehog* proteins which participate in protein-protein interactions involved in, for example, binding of the subject *hedgehog* polypeptide to other extracellular matrix components. To illustrate, the critical residues of a subject *hedgehog* polypeptide which are involved in molecular recognition of an *hedgehog* receptor such as *patched* can be determined and used to generate *hedgehog*-derived peptidomimetics which competitively inhibit binding of the authentic *hedgehog* protein with that moiety. By employing, for example, scanning mutagenesis to map the amino acid residues of each of the subject *hedgehog* proteins which are involved in binding other extracellular proteins, peptidomimetic compounds can be generated which mimic those residues of the *hedgehog* protein which facilitate the interaction. Such mimetics may then be used to interfere with the normal function of a *hedgehog* protein. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gama lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun*126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

Recombinantly produced forms of the hedgehog proteins can be produced using, e.g, expression vectors containing a nucleic acid encoding a *hedgehog* polypeptide, operably linked to at least one transcriptional regulatory sequence. Operably linked is intended to mean that the nucleotide sequence is linked to a regulatory sequence in a manner which allows expression of the nucleotide sequence. Regulatory sequences are art-recognized and are selected to direct expression of a *hedgehog* polypeptide. Accordingly, the term transcriptional regulatory sequence includes promoters, enhancers

and other expression control elements. Such regulatory sequences are described in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). For instance, any of a wide variety of expression control sequences, sequences that control the expression of a DNA sequence when operatively linked to it, may be used in these vectors to express DNA sequences encoding *hedgehog* polypeptide. Such useful expression control sequences, include, for example, a viral LTR, such as the LTR of the Moloney murine leukemia virus, the early and late promoters of SV40, adenovirus or cytomegalovirus immediate early promoter, the lac system, the trp system, the TAC or TRC system, T7 promoter whose expression is directed by T7 RNA polymerase, the major operator and promoter regions of phage λ, the control regions for fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase, e.g., Pho5, the promoters of the yeast α-mating factors, the polyhedron promoter of the baculovirus system and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof. It should be understood that the design of the expression vector may depend on such factors as the choice of the host cell to be transformed and/or the type of protein desired to be expressed. Moreover, the vector's copy number, the ability to control that copy number and the expression of any other proteins encoded by the vector, such as antibiotic markers, should also be considered.

In addition to providing a ready source of hedgehog polypeptides for purification, the gene constructs of the present invention can also be used as a part of a gene therapy protocol to deliver nucleic acids encoding either an agonistic or antagonistic form of a *hedgehog* polypeptide. Thus, another aspect of the invention features expression vectors for *in vivo* transfection of a *hedgehog* polypeptide in particular cell types so as cause ectopic expression of a *hedgehog* polypeptide in an muscle tissue.

Formulations of such expression constructs may be administered in any biologically effective carrier, e.g. any formulation or composition capable of effectively delivering the recombinant gene to cells *in vivo*. Approaches include insertion of the hedgehog coding sequence in viral vectors including recombinant retroviruses, adenovirus, adeno-associated virus, and herpes simplex virus-1, or recombinant bacterial or eukaryotic plasmids. Viral vectors transfect cells directly; plasmid DNA can be delivered with the help of, for example, cationic liposomes (lipofectin) or derivatized (e.g. antibody conjugated), polylysine conjugates, gramacidin S, artificial viral envelopes or other such intracellular carriers, as well as direct injection of the gene construct or $CaPO_4$ precipitation carried out *in vivo*. It will be appreciated that because transduction of appropriate target cells represents the critical first step in gene therapy, choice of the

particular gene delivery system will depend on such factors as the phenotype of the intended target and the route of administration, e.g. locally or systemically. Furthermore, it will be recognized that the particular gene construct provided for *in vivo* transduction of *hedgehog* expression are also useful for *in vitro* transduction of cells, such as for use in the *ex vivo* tissue culture systems described below.

A preferred approach for *in vivo* introduction of nucleic acid into a cell is by use of a viral vector containing nucleic acid, e.g. a cDNA, encoding the particular form of the *hedgehog* polypeptide desired. Infection of cells with a viral vector has the advantage that a large proportion of the targeted cells can receive the nucleic acid. Additionally, molecules encoded within the viral vector, e.g., by a cDNA contained in the viral vector, are expressed efficiently in cells which have taken up viral vector nucleic acid.

Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous genes *in vivo*, particularly into humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host. A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed "packaging cells") which produce only replication-defective retroviruses has increased the utility of retroviruses for gene therapy, and defective retroviruses are well characterized for use in gene transfer for gene therapy purposes (for a review see Miller, A.D. (1990) *Blood* 76:271). Thus, recombinant retrovirus can be constructed in which part of the retroviral coding sequence (*gag, pol, env*) has been replaced by nucleic acid encoding a hedgehog polypeptide and renders the retrovirus replication defective. The replication defective retrovirus is then packaged into virions which can be used to infect a target cell through the use of a helper virus by standard techniques. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* or *in vivo* with such viruses can be found in Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989), Sections 9.10-9.14 and other standard laboratory manuals. Examples of suitable retroviruses include pLJ, pZIP, pWE and pEM which are well known to those skilled in the art. Examples of suitable packaging virus lines for preparing both ecotropic and amphotropic retroviral systems include Crip, Cre, 2 and Am. Retroviruses have been used to introduce a variety of genes into many different cell types, including muscle cells, *in vitro* and/or *in vivo* (see for example Eglitis, et al. (1985) *Science* 230:1395-1398; Danos and Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:6460-6464; Wilson et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:3014-3018; Armentano et al. (1990) *Proc. Natl.*

Acad. Sci. USA 87:6141-6145; Huber et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8039-8043; Ferry et al. (1991) *Proc. Natl. Acad. Sci. USA* 88:8377-8381; Chowdhury et al. (1991) *Science* 254:1802-1805; van Beusechem et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:7640-7644; Kay et al. (1992) *Human Gene Therapy* 3:641-647; Dai et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:10892-10895; Hwu et al. (1993) *J. Immunol.* 150:4104-4115; U.S. Patent No. 4,868,116; U.S. Patent No. 4,980,286; PCT Application WO 89/07136; PCT Application WO 89/02468; PCT Application WO 89/05345; and PCT Application WO 92/07573).

Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle (see, for example PCT publications WO93/25234 and WO94/06920). For instance, strategies for the modification of the infection spectrum of retroviral vectors include: coupling antibodies specific for cell surface antigens to the viral *env* protein (Roux et al. (1989) *PNAS* 86:9079-9083; Julan et al. (1992) *J. Gen Virol* 73:3251-3255; and Goud et al. (1983) *Virology* 163:251-254); or coupling cell surface receptor ligands to the viral *env* proteins (Neda et al. (1991) *J Biol Chem* 266:14143-14146). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the *env* protein to an asialoglycoprotein), as well as by generating fusion proteins (e.g. single-chain antibody/*env* fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, can also be used to convert an ecotropic vector in to an amphotropic vector.

Moreover, use of retroviral gene delivery can be further enhanced by the use of tissue- or cell-specific transcriptional regulatory sequences which control expression of the *hedgehog* gene of the retroviral vector.

Another viral gene delivery system useful in the present method utilizes adenovirus-derived vectors. The genome of an adenovirus can be manipulated such that it encodes and expresses a gene product of interest but is inactivated in terms of its ability to replicate in a normal lytic viral life cycle. See for example Berkner et al. (1988) *BioTechniques* 6:616; Rosenfeld et al. (1991) *Science* 252:431-434; and Rosenfeld et al. (1992) *Cell* 68:143-155. Suitable adenoviral vectors derived from the adenovirus strain Ad type 5 dl324 or other strains of adenovirus (e.g., Ad2, Ad3, Ad7 etc.) are well known to those skilled in the art. Recombinant adenoviruses can be advantageous in certain circumstances in that they can be used to infect a wide variety of cell types, including muscle cells. Furthermore, the virus particle is relatively stable and amenable to purification and concentration, and as above, can be modified so as to affect

the spectrum of infectivity. Additionally, introduced adenoviral DNA (and foreign DNA contained therein) is not integrated into the genome of a host cell but remains episomal, thereby avoiding potential problems that can occur as a result of insertional mutagenesis in situations where introduced DNA becomes integrated into the host genome (e.g., retroviral DNA). Moreover, the carrying capacity of the adenoviral genome for foreign DNA is large (up to 8 kilobases) relative to other gene delivery vectors (Berkner et al. cited *supra*; Haj-Ahmand and Graham (1986) *J. Virol.* 57:267). Most replication-defective adenoviral vectors currently in use and therefore favored by the present invention are deleted for all or parts of the viral E1 and E3 genes but retain as much as 80% of the adenoviral genetic material (see, e.g., Jones et al. (1979) *Cell* 16:683; Berkner et al., *supra*; and Graham et al. in Methods in Molecular Biology, E.J. Murray, Ed. (Humana, Clifton, NJ, 1991) vol. 7. pp. 109-127). Expression of the inserted *hedgehog* gene can be under control of, for example, the E1A promoter, the major late promoter (MLP) and associated leader sequences, the E3 promoter, or exogenously added promoter sequences.

In addition to viral transfer methods, such as those illustrated above, non-viral methods can also be employed to cause expression of a *hedgehog* polypeptide in the tissue of an animal. Most nonviral methods of gene transfer rely on normal mechanisms used by mammalian cells for the uptake and intracellular transport of macromolecules. In preferred embodiments, non-viral gene delivery systems of the present invention rely on endocytic pathways for the uptake of the *hedgehog* polypeptide gene by the targeted cell. Exemplary gene delivery systems of this type include liposomal derived systems, poly-lysine conjugates, and artificial viral envelopes.

In clinical settings, the gene delivery systems for the therapeutic *hedgehog* gene can be introduced into a patient by any of a number of methods, each of which is familiar in the art. For instance, a pharmaceutical preparation of the gene delivery system can be introduced systemically, e.g. by intravenous injection, and specific transduction of the protein in the target cells occurs predominantly from specificity of transfection provided by the gene delivery vehicle, cell-type or tissue-type expression due to the transcriptional regulatory sequences controlling expression of the receptor gene, or a combination thereof. In other embodiments, initial delivery of the recombinant gene is more limited with introduction into the animal being quite localized. For example, the gene delivery vehicle can be introduced by catheter (see U.S. Patent 5,328,470) or by stereotactic injection (e.g. Chen et al. (1994) *PNAS* 91: 3054-3057). A *hedgehog* expression construct can be delivered in a gene therapy construct to dermal cells by, e.g.,

electroporation using techniques described, for example, by Dev et al. ((1994) *Cancer Treat Rev* 20:105-115).

The pharmaceutical preparation of the gene therapy construct can consist essentially of the gene delivery system in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery system can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can comprise one or more cells which produce the gene delivery system.

In yet another embodiment, the hedgehog or ptc therapeutic can be a "gene activation" construct which, by homologous recombination with a genomic DNA, alters the transcriptional regulatory sequences of an endogenous gene. For instance, the gene activation construct can replace the endogenous promoter of a *hedgehog* gene with a heterologous promoter, e.g., one which causes consitutive expression of the *hedgehog* gene or which causes inducible expression of the gene under conditions different from the normal expression pattern of the gene. Other genes in the *patched* signaling pathway can be similarly targeted. A vareity of different formats for the gene activation constructs are available. See, for example, the Transkaryotic Therapies, Inc PCT publications WO93/09222, WO95/31560, WO96/29411, WO95/31560 and WO94/12650.

In preferred embodiments, the nucleotide sequence used as the gene activation construct can be comprised of (1) DNA from some portion of the endogenous *hedgehog* gene (exon sequence, intron sequence, promoter sequences, etc.) which direct recombination and (2) heterologous transcriptional regulatory sequence(s) which is to be operably linked to the coding sequence for the genomic *hedgehog* gene upon recombination of the gene activation construct. For use in generating cultures of *hedgehog* producing cells, the construct may further include a reporter gene to detect the presence of the knockout construct in the cell.

The gene activation construct is inserted into a cell, and integrates with the genomic DNA of the cell in such a position so as to provide the heterologous regulatory sequences in operative association with the native *hedgehog* gene. Such insertion occurs by homologous recombination, i.e., recombination regions of the activation construct that are homologous to the endogenous *hedgehog* gene sequence hybridize to the genomic DNA and recombine with the genomic sequences so that the construct is incorporated into the corresponding position of the genomic DNA.

The terms "recombination region" or "targeting sequence" refer to a segment (i.e., a portion) of a gene activation construct having a sequence that is substantially identical

to or substantially complementary to a genomic gene sequence, e.g., including 5' flanking sequences of the genomic gene, and can facilitate homologous recombination between the genomic sequence and the targeting transgene construct.

As used herein, the term "replacement region" refers to a portion of a activation construct which becomes integrated into an endogenous chromosomal location following homologous recombination between a recombination region and a genomic sequence.

The heterologous regulatory sequences, e.g., which are provided in the replacement region, can include one or more of a variety elements, including: promoters (such as constitutive or inducible promoters), enhancers, negative regualtory elements, locus control regions, transcription factor binding sites, or combinations thereof.

Promoters/enhancers which may be used to control the expression of the targeted gene *in vivo* include, but are not limited to, the cytomegalovirus (CMV) promoter/enhancer (Karasuyama et al., 1989, *J. Exp. Med.*, 169:13), the human β-actin promoter (Gunning et al. (1987) *PNAS* 84:4831-4835), the glucocorticoid-inducible promoter present in the mouse mammary tumor virus long terminal repeat (MMTV LTR) (Klessig et al. (1984) *Mol. Cell Biol.* 4:1354-1362), the long terminal repeat sequences of Moloney murine leukemia virus (MuLV LTR) (Weiss et al. (1985) *RNA Tumor Viruses*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York), the SV40 early or late region promoter (Bernoist et al. (1981) *Nature* 290:304-310; Templeton et al. (1984) *Mol. Cell Biol.*, 4:817; and Sprague et al. (1983) *J. Virol.*, 45:773), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (RSV) (Yamamoto et al., 1980, *Cell*, 22:787-797), the herpes simplex virus (HSV) thymidine kinase promoter/enhancer (Wagner et al. (1981) *PNAS* 82:3567-71), and the herpes simplex virus LAT promoter (Wolfe et al. (1992) *Nature Genetics*, 1:379-384).

In an exemplary embodiment, portions of the 5' flanking region of the human Shh gene are amplified using primers which add restriction sites, to generate the following fragments

```
5'-
gcgcgcttcgaaGCGAGGCAGCCAGCGAGGGAGAGAGCGAGCGGGCGAGCCGGAGC-
GAGGAAatcgatgcgcgc (primer 1)
```

```
5'-
gcgcgcagatctGGGAAAGCGCAAGAGAGAGCGCACACGCACACACCCGCCGCGCG-
CACTCGggatccgcgcgc (primer 2)
```

As illustrated, primer 1 includes a 5' non-coding region of the human Shh gene and is flanked by an AsuII and ClaI restriction sites. Primer 2 includes a portion of the 5' non-coding region immediately 3' to that present in primer 1. The hedgehog gene sequence is flanked by XhoII and BamHI restriction sites. The purified amplimers are cut with each of the enzymes as appropriate.

The vector pCDNA1.1 (Invitrogen) includes a CMV promoter. The plasmid is cut with with AsuII, which cleaves just 3' to the CMV promoter sequence. The AsuII/ClaI fragment of primer 1 is ligated to the AsuII cleavage site of the pcDNA vector. The ClaI/AsuII ligation destroys the AsuII site at the 3' end of a properly inserted primer 1.

The vector is then cut with BamHI, and an XhoII/BamHI fragment of primer 2 is ligated to the BamHI cleavage site. As above, the BamHI/XhoII ligation destroys the BamHI site at the 5' end of a properly inserted primer 2.

Individual colonies are selected, cut with AsuII and BamHI, and the size of the AsuII/BamHI fragment determined. Colonies in which both the primer 1 and primer 2 sequences are corrcctly inserted are further amplified, an cut with AsuII and BamHI to produce the gene activation construct

```
cgaagcgaggcagccagcgagggagagagcgagcgggcgagccggagcgaggaaATCGA
AGGTTCGAATCCTTCCCCCACCACCATCACTTTCAAAAGTCCGAAAGAATCTGCTCCCT
GCTTGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGTAAAATTTAAGCTACAACAAGGCA
AGGCTTGACCGACAATTGCATGAAGAATCTGCTTAGGGTTAGGCGTTTTGCGCTGCTTC
GCGATGTACGGGCCAGATATACGCGTTGACATTGATTATTGACTAGTTATTAATAGTAA
TCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGA
CGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGACTAT
TTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCC
TATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTAT
GGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATG
CGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAG
TCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTC
CAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGG
GAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTAT
CGAAATTAATACGACTCACTATAGGGAGACCCAAGCTTGGTACCGAGCTCGGATCgatc
tgggaaagcgcaagagagagcgcacacgcacacacccgccgcgcactcgg
```

In this construct, the flanking primer 1 and primer 2 sequences provide the recombination region which permits the insertion of the CMV promoter in front of the coding sequence for the human *Shh* gene. Other heterologous promoters (or other transcriptional regulatory sequences) can be inserted in a genomic *hedgehog* gene by a similar method.

In still other embodiments, the replacement region merely deletes a negative transcriptional control element of the native gene, e.g., to activate expression, or ablates a positive control element, e.g., to inhibit expression of the targeted gene.

V. Exemplary ptc therapeutic compounds.

In another embodiment, the subject method is carried out using a ptc therapeutic composition. Such compositions can be generated with, for example, compounds which bind to patched and alter its signal transduction activity, compounds which alter the binding and/or enzymatic activity of a protein (e.g., intracellular) involved in patched signal pathway, and compounds which alter the level of expression of a hedgehog protein, a patched protein or a protein involved in the intracellular signal transduction pathway of patched.

The availability of purified and recombinant *hedgehog* polypeptides facilitates the generation of assay systems which can be used to screen for drugs, such as small organic molecules, which are either agonists or antagonists of the normal cellular function of a *hedgehog* and/or patched protein, particularly their role in the pathogenesis of muscle cell proliferation and/or differentiation. In one embodiment, the assay evaluates the ability of a compound to modulate binding between a *hedgehog* polypeptide and a *hedgehog* receptor such as *patched*. In other embodiments, the assay merely scores for the ability of a test compound to alter the signal transduction acitity of the *patched* protein. In this manner, a variety of *hedgehog* and/or *ptc* therapeutics, both proliferative and anti-proliferative in activity, can be identified. A variety of assay formats will suffice and, in light of the present disclosure, will be comprehended by skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test

compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with receptor proteins.

Acordingly, in an exemplary screening assay for *ptc* therapeutics, the compound of interest is contacted with a mixture including a *hedgehog* receptor protein (e.g., a cell expressing the *patched* receptor) and a hedgehog protein under conditions in which it is ordinarily capable of binding the *hedgehog* protein. To the mixture is then added a composition containing a test compound. Detection and quantification of receptor/*hedgehog* complexes provides a means for determining the test compound's efficacy at inhibiting (or potentiating) complex formation between the receptor protein and the *hedgehog* polypeptide. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. In the control assay, isolated and purified *hedgehog* polypeptide is added to the receptor protein, and the formation of receptor/*hedgehog* complex is quantitated in the absence of the test compound.

In other embodiments, a ptc therapeutic of the present invention is one which disrupts the association of *patched* with *smoothened.*

Agonist and antagonists of muscle cell growth can be distinguished, and the efficacy of the compound can be assessed, by subsequent testing with muscle cells, e.g., in culture.

In an illustrative embodiment, the polypeptide utilized as a *hedgehog* receptor can be generated from the *patched* protein. Accordingly, an exemplary screening assay includes all or a suitable portion of the *patched* protein which can be obtained from, for example, the human *patched* gene (GenBank U43148) or other vertebrate sources (see GenBank Accession numbers U40074 for chicken *patched* and U46155 for mouse *patched*), as well as from drosophila (GenBank Accession number M28999) or other invertebrate sources. The *patched* protein can be provided in the screening assay as a whole protein (preferably expressed on the surface of a cell), or alternatively as a fragment of the full length protein which binds to *hedgehog* polypeptides, e.g., as one or both of the substantial extracellular domains (e.g. corresponding to residues Asn120-Ser438 and/or Arg770-Trp1027 of the human *patched* protein - which are also potential antagonists of *hedgehog*-dependent signal transduction). For instance, the *patched* protein can be provided in soluble form, as for example a preparation of one of the extracellular domains, or a preparation of both of the extracellular domains which are

covalently connected by an unstructured linker (see, for example, Huston et al. (1988) PNAS 85:4879; and U.S. Patent No. 5,091,513). In other embodiments, the protein can be provided as part of a liposomal preparation or expressed on the surface of a cell. The *patched* protein can derived from a recombinant gene, e.g., being ectopically expressed in a heterologous cell. For instance, the protein can be expressed on oocytes, mammalian cells (e.g., COS, CHO, 3T3 or the like), or yeast cell by standard recombinant DNA techniques. These recombinant cells can be used for receptor binding, signal transduction or gene expression assays. Marigo et al. (1996) *Development* 122:1225-1233 illustrates a binding assay of human *hedgehog* to chick *patched* protein ectopically expressed in *Xenopus laevis* oocytes. The assay system of Marigo et al. can be adapted to the present drug screening assays. As illustrated in that reference, *Shh* binds to the *patched* protein in a selective, saturable, dose-dependent manner, thus demonstrating that *patched* is a receptor for *Shh*.

Complex formation between the *hedgehog* polypeptide and a *hedgehog* receptor may be detected by a variety of techniques. For instance, modulation of the formation of complexes can be quantitated using, for example, detectably labelled proteins such as radiolabelled, fluorescently labelled, or enzymatically labelled *hedgehog* polypeptides, by immunoassay, or by chromatographic detection.

Typically, for cell-free assays, it will be desirable to immobilize either the *hedgehog* receptor or the *hedgehog* polypeptide to facilitate separation of receptor/*hedgehog* complexes from uncomplexed forms of one of the proteins, as well as to accommodate automation of the assay. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/receptor (GST/receptor) fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the *hedgehog* polypeptide, e.g. an ^{35}S-labeled *hedgehog* polypeptide, and the test compound and incubated under conditions conducive to complex formation, e.g. at physiological conditions for salt and pH, though slightly more stringent conditions may be desired. Following incubation, the beads are washed to remove any unbound *hedgehog* polypeptide, and the matrix bead-bound radiolabel determined directly (e.g. beads placed in scintillant), or in the supernatant after the receptor/*hedgehog* complexes are dissociated. Alternatively, the complexes can be dissociated from the bead, separated by SDS-PAGE gel, and the level of *hedgehog* polypeptide found in the bead fraction quantitated from the gel using standard electrophoretic techniques.

Other techniques for immobilizing proteins on matrices are also available for use in the subject assay. For instance, soluble portions of the *hedgehog* receptor protein can be immobilized utilizing conjugation of biotin and streptavidin. For instance, biotinylated receptor molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the *hedgehog* receptor but which do not interfere with *hedgehog* binding can be derivatized to the wells of the plate, and the receptor trapped in the wells by antibody conjugation. As above, preparations of a *hedgehog* polypeptide and a test compound are incubated in the receptor-presenting wells of the plate, and the amount of receptor/*hedgehog* complex trapped in the well can be quantitated. Exemplary methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the *hedgehog* polypeptide, or which are reactive with the receptor protein and compete for binding with the *hedgehog* polypeptide; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the *hedgehog* polypeptide. In the instance of the latter, the enzyme can be chemically conjugated or provided as a fusion protein with the *hedgehog* polypeptide. To illustrate, the *hedgehog* polypeptide can be chemically cross-linked or genetically fused with alkaline phosphatase, and the amount of *hedgehog* polypeptide trapped in the complex can be assessed with a chromogenic substrate of the enzyme, e.g. paranitrophenylphosphate. Likewise, a fusion protein comprising the *hedgehog* polypeptide and glutathione-S-transferase can be provided, and complex formation quantitated by detecting the GST activity using 1-chloro-2,4-dinitrobenzene (Habig et al (1974) *J Biol Chem* 249:7130).

For processes which rely on immunodetection for quantitating one of the proteins trapped in the complex, antibodies against the protein, such as the anti-*hedgehog* antibodies described herein, can be used. Alternatively, the protein to be detected in the complex can be "epitope tagged" in the form of a fusion protein which includes, in addition to the *hedgehog* polypeptide or *hedgehog* receptor sequence, a second polypeptide for which antibodies are readily available (e.g. from commercial sources). For instance, the GST fusion proteins described above can also be used for quantification of binding using antibodies against the GST moiety. Other useful epitope tags include myc-epitopes (e.g., see Ellison et al. (1991) *J Biol Chem* 266:21150-21157) which includes a 10-residue sequence from c-myc, as well as the pFLAG system (International Biotechnologies, Inc.) or the pEZZ-protein A system (Pharamacia, NJ).

Where the desired portion of the *hedgehog* receptor (or other *hedgehog* binding molecule) cannot be provided in soluble form, liposomal vesicles can be used to provide manipulatable and isolatable sources of the receptor. For example, both authentic and recombinant forms of the *patched* protein can be reconstituted in artificial lipid vesicles (e.g. phosphatidylcholine liposomes) or in cell membrane-derived vesicles (see, for example, Bear et al. (1992) *Cell* 68:809-818; Newton et al. (1983) *Biochemistry* 22:6110-6117; and Reber et al. (1987) *J Biol Chem* 262:11369-11374).

In addition to cell-free assays, such as described above, the readily available source of *hedgehog* proteins provided by the art also facilitates the generation of cell-based assays for identifying small molecule agonists/antagonists and the like. Analogous to the cell-based assays described above for screening combinatorial libraries, cells which are sensitive to *hedgehog* induction, e.g. *patched*-expressing cells or other myoblast-derived cells sensitive to *hedgehog* induction, can be contacted with a *hedgehog* protein and a test agent of interest, with the assay scoring for anything from simple binding to the cell to modulation in *hedgehog* inductive responses by the target cell in the presence and absence of the test agent. As with the cell-free assays, agents which produce a statistically significant change in *hedgehog* activities (either inhibition or potentiation) can be identified.

In other emdodiments, the cell-based assay scores for agents which disrupt association of patched and *smoothened* proteins, e.g., in the cell surface membrane or liposomal preparation.

In addition to characterizing cells that naturally express the *patched* protein, cells which have been genetically engineered to ectopically express *patched* can be utilized for drug screening assays. As an example, cells which either express low levels or lack expression of the *patched* protein, e.g. *Xenopus laevis* oocytes, COS cells or yeast cells, can be genetically modified using standard techniques to ectopically express the *patched* protein. (see Marigo et al., *supra*).

The resulting recombinant cells, e.g., which express a functional *patched* receptor, can be utilized in receptor binding assays to identify agonist or anatagonsts of *hedgehog* binding. Binding assays can be performed using whole cells. Furthermore, the recombinant cells of the present invention can be engineered to include other heterolgous genes encoding proteins involved in *hedgehog*-dependent siganl pathways. For example, the gene products of one or more of *smoothened*, *costal-2* and/or *fused* can be co-expressed with *patched* in the reagent cell, with assays being sensitive to the functional reconstituion of the *hedgehog* signal transduction cascade.

Alternatively, liposomal preparations using reconstituted *patched* protein can be utilized. *Patched* protein purified from detergent extracts from both authentic and recombinant origins can be reconstituted in in artificial lipid vesicles (e.g. phosphatidylcholine liposomes) or in cell membrane-derived vesicles (see, for example, Bear et al. (1992) *Cell* 68:809-818; Newton et al. (1983) *Biochemistry* 22:6110-6117; and Reber et al. (1987) *J Biol Chem* 262:11369-11374). The lamellar structure and size of the resulting liposomes can be characterized using electron microscopy. External orientation of the *patched* protein in the reconstituted membranes can be demonstrated, for example, by immunoelectron microscopy. The *hedgehog* protein binding activity of liposomes containing *patched* and liposomes without the protein in the presence of candidate agents can be compared in order to identify potential modulators of the *hedgehog-patched* interaction.

The *hedgehog* protein used in these cell-based assays can be provided as a purified source (natural or recombinant in origin), or in the form of cells/tissue which express the protein and which are co-cultured with the target cells. As in the cell-free assays, where simple binding (rather than induction) is the hedgehog activity scored for in the assay, the protein can bc labelled by any of the above-mentioned techniques, e.g., fluorescently, enzymatically or radioactively, or detected by immunoassay.

In addition to binding studies, functional assays can be used to identified modulators, i.e., agonists or antagonists, of *hedgehog* or *patched* activities. By detecting changes in intracellular signals, such as alterations in second messengers or gene expression, in *patched*-expressing cells contacted with a test agent, candidate agonists and antagonists to *patched* signaling can be identified.

A number of gene products have been implicated in *patched*-mediated signal transduction, including *patched*, the transcription factor *cubitus interruptus* (ci), the serine/threonine kinase *fused* (fu) and the gene products of *costal-2*, *smoothened* and *suppressor of fused*.

The interaction of a hedgehog protein with *patched* sets in motion a cascade involving the activation and inhibition of downstream effectors, the ultimate consequence of which is, in some instances, a detectable change in the transcription or translation of a gene. Potential transcriptional targets of *patched* signaling are the *patched* gene itself (Hidalgo and Ingham, 1990 *Development* 110, 291-301; Marigo et al., 1996) and the vertebrate homologs of the drosophila cubitus interruptus gene, the *GLI* genes (Hui et al. (1994) *Dev Biol* 162:402-413). *Patched* gene expression has been shown to be induced in cells of the limb bud and the neural plate that are responsive to *Shh*. (Marigo et al. (1996) *PNAS*, in press; Marigo et al. (1996) *Development* 122:1225-

1233). The *GLI* genes encode putative transcription factors having zinc finger DNA binding domains (Orenic et al. (1990) *Genes & Dev* 4:1053-1067; Kinzler et al. (1990) *Mol Cell Biol* 10:634-642). Transcription of the *GLI* gene has been reported to be upregulated in response to *hedgehog* in limb buds, while transcription of the *GLI3* gene is downregulated in response to *hedgehog* induction (Marigo et al. (1996) *Development* 122:1225-1233). By selecting transcriptional regulatory sequences from such target genes, e.g. from *patched* or *GLI* genes, that are responsible for the up- or down regulation of these genes in response to *patched* signalling, and operatively linking such promoters to a reporter gene, one can derive a transcription based assay which is sensitive to the ability of a specific test compound to modify *patched* signalling pathways. Expression of the reporter gene, thus, provides a valuable screening tool for the development of compounds that act as agonists or antagonists of *ptc* induction of differentiation/quiescence.

Reporter gene based assays of this invention measure the end stage of the above described cascade of events, e.g., transcriptional modulation. Accordingly, in practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on *ptc* signaling. To identify potential regulatory elements responsive to *ptc* signaling present in the transcriptional regulatory sequence of a target gene, nested deletions of genomic clones of the target gene can be constructed using standard techniques. See, for example, Current Protocols in Molecular Biology, Ausubel, F.M. et al. (eds.) Greene Publishing Associates, (1989); U.S. Patent 5,266,488; Sato et al. (1995) *J Biol Chem* 270:10314-10322; and Kube et al. (1995) *Cytokine* 7:1-7. A nested set of DNA fragments from the gene's 5'-flanking region are placed upstream of a reporter gene, such as the luciferase gene, and assayed for their ability to direct reporter gene expression in *patched* expressing cells. Host cells transiently transfected with reporter gene constructs can be scored for the induction of expression of the reporter gene in the presence and absence of *hedgehog* to determine regulatory sequences which are responsice to *patched*-dependent signalling.

In practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on second messengers generated by induction with *hedgehog* protein. Typically, the reporter gene construct will include a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the *hedgehog* activity, with the level of expression of the reporter gene providing the *hedgehog*-dependent detection signal. The amount of transcription from the reporter gene may be measured using any method known to those of skill in the art to be suitable. For example, mRNA expression from the

reporter gene may be detected using RNAse protection or RNA-based PCR, or the protein product of the reporter gene may be identified by a characteristic stain or an intrinsic activity. The amount of expression from the reporter gene is then compared to the amount of expression in either the same cell in the absence of the test compound (or *hedgehog*) or it may be compared with the amount of transcription in a substantially identical cell that lacks the target receptor protein. Any statistically or otherwise significant difference in the amount of transcription indicates that the test compound has in some manner altered the signal transduction of the *patched* protein, e.g., the test compound is a potential ptc therapeutic.

As described in further detail below, in preferred embodiments the gene product of the reporter is detected by an intrinsic activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detection signal based on color, fluorescence, or luminescence. In other preferred embodiments, the reporter or marker gene provides a selective growth advantage, e.g., the reporter gene may enhance cell viability, relieve a cell nutritional requirement, and/or provide resistance to a drug.

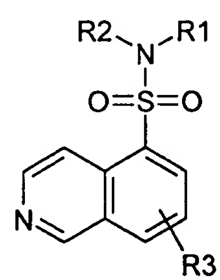
Preferred reporter genes are those that are readily detectable. The reporter gene may also be included in the construct in the form of a fusion gene with a gene that includes desired transcriptional regulatory sequences or exhibits other desirable properties. Examples of reporter genes include, but are not limited to CAT (chloramphenicol acetyl transferase) (Alton and Vapnek (1979), Nature 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet et al. (1987), Mol. Cell. Biol. 7:725-737); bacterial luciferase (Engebrecht and Silverman (1984), PNAS 1: 4154-4158; Baldwin et al. (1984), Biochemistry 23: 3663-3667); alkaline phosphatase (Toh et al. (1989) Eur. J. Biochem. 182: 231-238, Hall et al. (1983) J. Mol. Appl. Gen. 2: 101), human placental secreted alkaline phosphatase (Cullen and Malim (1992) Methods in Enzymol. 216:362-368).

Transcriptional control elements which may be included in a reporter gene construct include, but are not limited to, promoters, enhancers, and repressor and activator binding sites. Suitable transcriptional regulatory elements may be derived from the transcriptional regulatory regions of genes whose expression is induced after modulation of a *patched* signal transduction pathway. The characteristics of preferred genes from which the transcriptional control elements are derived include, but are not limited to, low or undetectable expression in quiescent cells, rapid induction at the transcriptional level within minutes of extracellular simulation, induction that is transient and independent of new protein synthesis, subsequent shut-off of transcription requires

new protein synthesis, and mRNAs transcribed from these genes have a short half-life. It is not necessary for all of these properties to be present.

In yet other embodiments, second messenger generation can be measured directly in the detection step, such as mobilization of intracellular calcium, phospholipid metabolism or adenylate cyclase activity are quantitated, for instance, the products of phospholipid hydrolysis IP_3, DAG or cAMP could be measured For example, recent studies have implicated protein kinase A (PKA) as a possible component of *hedgehog/patched* signaling (Hammerschmidt et al. (1996) *Genes & Dev* 10:647). High PKA activity has been shown to antagonize *hedgehog* signaling in these systems. Although it is unclear whether PKA acts directly downstream or in parallel with *hedgehog* signaling, it is possible that *hedgehog* signalling occurs via inhibition of PKA activity. Thus, detection of PKA activity provides a potential readout for the instant assays.

In a preferred embodiment, the *ptc* therapeutic is a PKA inhibitor. A variety of PKA inhibitors are known in the art, including both peptidyl and organic compounds. For instance, the *ptc* therapeutic can be a 5-isoquinolinesulfonamide, such as represented in the general formula:



wherein,

R_1 and R_2 each can independently represent hydrogen, and as valence and stability permit a lower alkyl, a lower alkenyl, a lower alkynyl, a carbonyl (such as a carboxyl, an ester, a formate, or a ketone), a thiocarbonyl (such as a thioester, a thioacetate, or a thioformate), an amino, an acylamino, an amido, a cyano, a nitro, an azido, a sulfate, a sulfonate, a sulfonamido, $-(CH_2)_m-R_8$, $-(CH_2)_m-OH$, $-(CH_2)_m$-O-lower alkyl, $-(CH_2)_m$-O-lower alkenyl, $-(CH_2)_n-O-(CH_2)_m-R_8$, $-(CH_2)_m-SH$, $-(CH_2)_m$-S-lower alkyl, $-(CH_2)_m$-S-lower alkenyl, $-(CH_2)_n-S-(CH_2)_m-R_8$, or

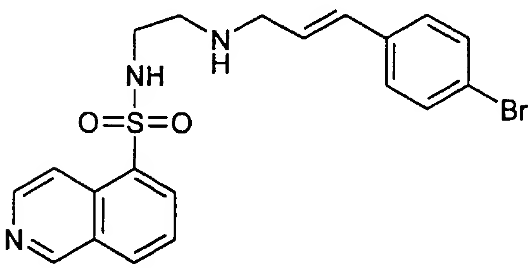
R_1 and R_2 taken together with N form a heterocycle (substituted or unsubstituted);

R_3 is absent or represents one or more substitutions to the isoquinoline ring such as a lower alkyl, a lower alkenyl, a lower alkynyl, a carbonyl (such as a carboxyl, an ester, a formate, or a ketone), a thiocarbonyl (such as a thioester, a thioacetate, or a thioformate), an amino, an acylamino, an amido, a cyano, a nitro, an azido, a sulfate, a sulfonate, a sulfonamido, $-(CH_2)_m-R_8$, $-(CH_2)_m-OH$, $-(CH_2)_m$-O-lower alkyl, $-(CH_2)_m$-O-lower alkenyl, $-(CH_2)_n-O-(CH_2)_m-R_8$, $-(CH_2)_m-SH$, $-(CH_2)_m$-S-lower alkyl, $-(CH_2)_m$-S-lower alkenyl, $-(CH_2)_n-S-(CH_2)_m-R_8$;

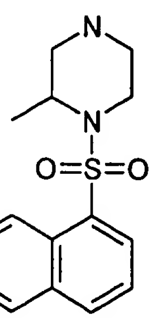
R_8 represents a substituted or unsubstituted aryl, aralkyl, cycloalkyl, cycloalkenyl, or heterocycle; and

n and m are independently for each occurrence zero or an integer in the range of 1 to 6.

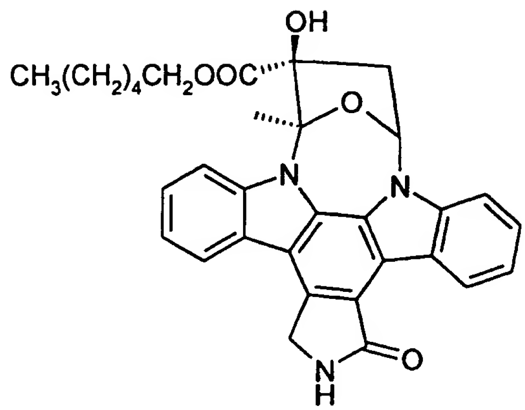
In a preferred embodiment, the PKA inhibitor is N-[2-((p-bromocinnamyl)amino)ethyl]-5-isoquinolinesulfonamide (H-89; Calbiochem Cat. No. 371963), e.g., having the formula:



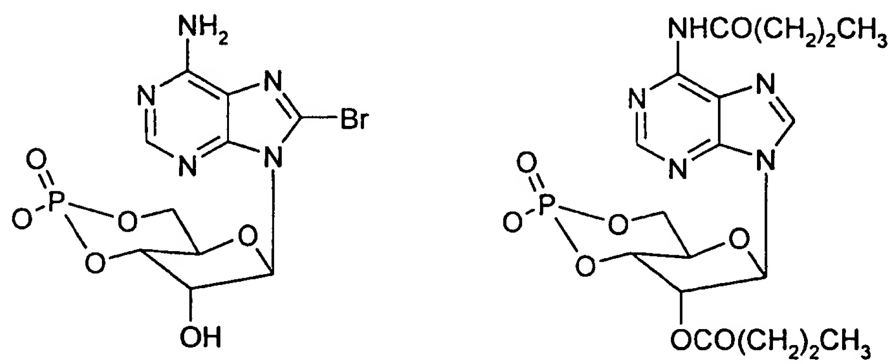
In another embodiment, the PKA inhibitor is 1-(5-isoquinolinesulfonyl)-2-methylpiperazine (H-7; Calbiochem Cat. No. 371955), e.g., having the formula:



In still other embodiments, the PKA inhibitor is KT5720 (Calbiochem Cat. No. 420315), having the structure



A variety of nucleoside analogs are also useful as PKA inhibitors. For example, the subject method can be carried out cyclic AMP analogs which inhibit the kinase activity of PKA, as for example, 8-bromo-cAMP or dibutyryl-cAMP



Exemplary peptidyl inhibitors of PKA activity include the PKA Heat Stable Inhibitor (isoform α; see, for example, Calbiochem Cat. No. 539488, and Wen et al. (1995) *J Biol Chem* 270:2041).

Certain *hedehog* receptors may stimulate the activity of phospholipases. Inositol lipids can be extracted and analyzed using standard lipid extraction techniques. Water soluble derivatives of all three inositol lipids (IP_1, IP_2, IP_3) can also be quantitated using radiolabelling techniques or HPLC.

The mobilization of intracellular calcium or the influx of calcium from outside the cell may be a response to *hedgehog* stimulation or lack there of. Calcium flux in the reagent cell can be measured using standard techniques. The choice of the appropriate calcium indicator, fluorescent, bioluminescent, metallochromic, or Ca^{++}-sensitive microelectrodes depends on the cell type and the magnitude and time constant of the

event under study (Borle (1990) *Environ Health Perspect* 84:45-56). As an exemplary method of Ca^{++} detection, cells could be loaded with the Ca^{++}sensitive fluorescent dye fura-2 or indo-1, using standard methods, and any change in Ca^{++} measured using a fluorometer.

In certain embodiments of the assay, it may be desirable to screen for changes in cellular phosphorylation. As an example, the drosophila gene *fused* (fu) which encodes a serine/threonine kinase has been identified as a potential downstream target in *hedgehog* signaling. (Preat et al., 1990 *Nature* 347, 87-89; Therond et al. 1993, *Mech. Dev.* 44. 65-80). The ability of compounds to modulate serine/threonine kinase activation could be screened using colony immunoblotting (Lyons and Nelson (1984) *Proc. Natl. Acad. Sci. USA* 81:7426-7430) using antibodies against phosphorylated serine or threonine residues. Reagents for performing such assays are commercially available, for example, phosphoserine and phosphothreonine specific antibodies which measure increases in phosphorylation of those residues can be purchased from comercial sources.

In yet another embodiment, the *ptc* therapeutic is an antisense molecule which inhibits expression of a protein involved in a *patched*-mediated signal transduction pathway. To illustrate, by inhibiting the expression of a protein which are involved in *patched* signals, such as fused, costal-2, smoothened and/or Gli genes, the ability of the patched signal pathway(s) to inhibit proliferation of a cell can be altered, e.g., potentiated or repressed.

As used herein, "antisense" therapy refers to administration or *in situ* generation of oligonucleotide probes or their derivatives which specifically hybridize (e.g. bind) under cellular conditions with cellular mRNA and/or genomic DNA encoding a *hedgehog* protein, patched, or a protein involved in patched-mediated signal transduction. The hybridization should inhibit expression of that protein, e.g. by inhibiting transcription and/or translation. The binding may be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interactions in the major groove of the double helix. In general, "antisense" therapy refers to the range of techniques generally employed in the art, and includes any therapy which relies on specific binding to oligonucleotide sequences.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid which, when transcribed in the cell, produces RNA which is complementary to at least a unique portion of the target cellular mRNA. Alternatively, the antisense construct is an oligonucleotide probe which is generated *ex vivo* and which, when introduced into the cell causes inhibition of expression by hybridizing with the

mRNA and/or genomic sequences of a target gene. Such oligonucleotide probes are preferably modified oligonucleotide which are resistant to endogenous nucleases, e.g. exonucleases and/or endonucleases, and is therefore stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patents 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy have been reviewed, for example, by Van der Krol et al. (1988) *Biotechniques* 6:958-976; and Stein et al. (1988) *Cancer Res* 48:2659-2668.

Several considerations should be taken into account when constructing antisense oligonucleotides for the use in the methods of the invention: (1) oligos should have a GC content of 50% or more; (2) avoid sequences with stretches of 3 or more G's; and (3) oligonucleotides should not be longer than 25-26 mers. When testing an antisense oligonucleotide, a mismatched control can be constructed. The controls can be generated by reversing the sequence order of the corresponding antisense oligonucleotide in order to conserve the same ratio of bases.

In an illustrative embodiment, the *ptc* therapeutic can be an antisense construct for inhibiting the expression of *patched*, e.g., to mimic the inhibition of *patched* by *hedgehog*. Exemplary antisense constructs include:

5'-GTCCTGGCGCCGCCGCCGCCGTCGCC

5'-TTCCGATGACCGGCCTTTCGCGGTGA

5'-GTGCACGGAAAGGTGCAGGCCACACT

VI. Exemplary pharmaceutical preparations of hedgehog and ptc therapeutics

The source of the hedgehog and ptc therapeutics to be formulated will depend on the particular form of the agent. Small organic molecules and peptidyl fragments can be chemically synthesized and provided in a pure form suitable for pharmaceutical/cosmetic usage. Products of natural extracts can be purified according to techniques known in the art. For example, the Cox et al. U.S. Patent 5,286,654 describes a method for purifying naturally occurring forms of a secreted protein and can be adapted for purification of hedgehog polypeptides. Recombinant sources of hedgehog polypeptides are also available. For example, the gene encoding *hedgehog* polypeptides, are known, *inter alia*, from PCT publications WO 95/18856 and WO 96/17924.

Those of skill in treating muscle tissues can determine the effective amount of an hedgehog or ptc therapeutic to be formulated in a pharmaceutical or cosmetic preparation.

The hedgehog or ptc therapeutic formulations used in the method of the invention are most preferably applied in the form of appropriate compositions. As appropriate compositions there may be cited all compositions usually employed for systemically or topically administering drugs. The pharmaceutically acceptable carrier should be substantially inert, so as not to act with the active component. Suitable inert carriers include water, alcohol polyethylene glycol, mineral oil or petroleum gel, propylene glycol and the like.

To prepare the pharmaceutical compositions of this invention, an effective amount of the particular hedgehog or ptc therapeutic as the active ingredient is combined in intimate admixture with a pharmaceutically acceptable carrier, which carrier may take a wide variety of forms depending on the form of preparation desired for administration. These pharmaceutical compositions are desirable in unitary dosage form suitable, particularly, for administration orally, rectally, percutaneously, or by parenteral injection. For example, in preparing the compositions in oral dosage form, any of the usual pharmaceutical media may be employed such as, for example, water, glycols, oils, alcohols and the like in the case of oral liquid preparations such as suspensions, syrups, elixirs and solutions; or solid carriers such as starches, sugars, kaolin, lubricants, binders, disintegrating agents and the like in the case of powders, pills, capsules, and tablets. Because of their ease in administration, tablets and capsules represents the most advantageous oral dosage unit form, in which case solid pharmaceutical carriers are obviously employed. For parenteral compositions, the carrier will usually comprise sterile water, at least in large part, though other ingredients, for example, to aid solubility, may be included. Injectable solutions, for example, may be prepared in which the carrier comprises saline solution, glucose solution or a mixture of saline and glucose solution. Injectable suspensions may also be prepared in which case appropriate liquid carriers, suspending agents and the like may be employed. Also included are solid form preparations which are intended to be converted, shortly before use, to liquid form preparations. In the compositons suitable for percutaneous administration, the carrier optionally comprises a penetration enhancing agent and/or a suitable wetting agent, optionally combined with suitable additives of any nature in minor proportions, which additives do not introduce a significant deleterious effect on the skin.

In addition to the direct topical application of the preparations they can be topically administered by other methods, for example, encapsulated in a temperature and/or

pressure sensitive matrix or in film or solid carrier which is soluble in body fluids and the like for subsequent release, preferably sustained-release of the active component.

As appropriate compositions for topical application there may be cited all compositions usually employed for topically administering therapeuitcs, e.g., creams, gellies, dressings, shampoos, tinctures, pastes, ointments, salves, powders, liquid or semiliquid formulation and the like. Application of said compositions may be by aerosol e.g. with a propellent such as nitrogen carbon dioxide, a freon, or without a propellent such as a pump spray, drops, lotions, or a semisolid such as a thickened composition which can be applied by a swab. In particular compositions, semisolid compositions such as salves, creams, pastes, gellies, ointments and the like will conveniently be used.

It is especially advantageous to formulate the subject compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used in the specification and claims herein refers to physically discreate units suitable as unitary dosages, each unit containing a predetermined quantity of active ingredient calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. Examples of such dosage unit forms are tablets (including scored or coated tablets), capsules, pills, powders packets, wafers, injectable solutions or suspensions, teaspoonfuls, tablespoonfuls and the like, and segregated multiples thereof.

The pharmaceutical preparations of the present invention can be used, as stated above, for the many applications whcih can be considered cosmetic uses. Cosmetic compositions known in the art, preferably hypoallergic and pH controlled are especially preferred, and include toilet waters, packs, lotions, skin milks or milky lotions. The preparations contain, besides the hedgehog or ptc therapeutic, components usually employed in such preparations. Examples of such components are oils, fats, waxes, surfactants, humectants, thickening agents, antioxidants, viscosity stabilizers, chelating agents, buffers, preservatives, perfumes, dyestuffs, lower alkanols, and the like. If desired, further ingredients may be incorporated in the compositions, e.g. antiinflammatory agents, antibacterials, antifungals, disinfectants, vitamins, sunscreens, antibiotics, or other anti-acne agents.

Examples of oils comprise fats and oils such as olive oil and hydrogenated oils; waxes such as beeswax and lanolin; hydrocarbons such as liquid paraffin, ceresin, and squalane; fatty acids such as stearic acid and oleic acid; alcohols such as cetyl alcohol, stearyl alcohol, lanolin alcohol, and hexadecanol; and esters such as isopropyl myristate, isopropyl palmitate and butyl stearate. As examples of surfactants there may be cited anionic surfactants such as sodium stearate, sodium cetylsulfate, polyoxyethylene

laurylether phosphate, sodium N-acyl glutamate; cationic surfactants such as stearyldimethylbenzylammonium chloride and stearyltrimethylammonium chloride; ampholytic surfactants such as alkylaminoethylglycine hydrocloride solutions and lecithin; and nonionic surfactants such as glycerin monostearate, sorbitan monostearate, sucrose fatty acid esters, propylene glycol monostearate, polyoxyethylene oleylether, polyethylene glycol monostearate, polyoxyethylene sorbitan monopalmitate, polyoxyethylene coconut fatty acid monoethanolamide, polyoxypropylene glycol (e.g. the materials sold under the trademark "Pluronic"), polyoxyethylene castor oil, and polyoxyethylene lanolin. Examples of humectants include glycerin, 1,3-butylene glycol, and propylene glycol; examples of lower alcohols include ethanol and isopropanol; examples of thickening agents include xanthan gum, hydroxypropyl cellulose, hydroxypropyl methyl cellulose, polyethylene glycol and sodium carboxymethyl cellulose; examples of antioxidants comprise butylated hydroxytoluene, butylated hydroxyanisole, propyl gallate, citric acid and ethoxyquin; examples of chelating agents include disodium edetate and ethanehydroxy diphosphate; examples of buffers comprise citric acid, sodium citrate, boric acid, borax, and disodium hydrogen phosphate; and examples of preservatives are methyl parahydroxybenzoate, ethyl parahydroxybenzoate, dehydroacetic acid, salicylic acid and benzoic acid.

For preparing ointments, creams, toilet waters, skin milks, and the like, typically from 0.01 to 10% in particular from 0.1 to 5% and more in particular from 0.2 to 2.5% of the active ingredient, e.g., of the hedgehog or ptc therapeutic, will be incorporated in the compositions. In ointments or creams, the carrier for example consists of 1 to 20%, in particular 5 to 15% of a humectant, 0.1 to 10% in particular from 0.5 to 5% of a thickener and water; or said carrier may consist of 70 to 99%, in particular 20 to 95% of a surfactant, and 0 to 20%, in particular 2.5 to 15% of a fat; or 80 to 99.9% in particular 90 to 99% of a thickener; or 5 to 15% of a surfactant, 2-15% of a humectant, 0 to 80% of an oil, very small (< 2%) amounts of preservative, coloring agent and/or perfume, and water. In a toilet water, the carrier for example consists of 2 to 10% of a lower alcohol, 0.1 to 10% or in particular 0.5 to 1% of a surfactant, 1 to 20%, in particular 3 to 7% of a humectant, 0 to 5% of a buffer, water and small amounts (< 2%) of preservative, dyestuff and/or perfume. In a skin milk, the carrier typically consists of 10-50% of oil, 1 to 10% of surfactant, 50-80% of water and 0 to 3% of preservative and/or perfume. In the aforementioned preparations, all % symbols refer to weight by weight percentage.

Particular compositions for use in the method of the present invention are those wherein the hedgehog or ptc therapeutic is formulated in liposome-containing compositions. Liposomes are artificial vesicles formed by amphiphatic molecules such as

polar lipids, for example, phosphatidyl cholines, ethanolamines and serines, sphingomyelins, cardiolipins, plasmalogens, phosphatidic acids and cerebiosides. Liposomes are formed when suitable amphiphathic molecules are allowed to swell in water or aqueous solutions to form liquid crystals usually of multilayer structure comprised of many bilayers separated from each other by aqueous material (also referred to as coarse liposomes). Another type of liposome known to be consisting of a single bilayer encapsulating aqueous material is referred to as a unilamellar vesicle. If water-soluble materials are included in the aqueous phase during the swelling of the lipids they become entrapped in the aqueous layer between the lipid bilayers.

Water-soluble active ingredients such as, for example, various salt forms of a hedgehog polypeptide, are encapsulated in the aqueous spaces between the molecular layers. The lipid soluble active ingredient of hedgehog or ptc therapeutic, such as an organic mimetic, is predominantly incorporated into the lipid layers, although polar head groups may protude from the layer into the aqueous space. The encapsulation of these compounds can be achieved by a number of methods. The method most commonly used involves casting a thin film of phospholipid onto the walls of a flask by evaporation from an organic solvent. When this film is dispersed in a suitable aqueous medium, multilamellar liposomes are formed. Upon suitable sonication, the coarse liposomes form smaller similarly closed vesicles.

Water-soluble active ingredients are usually incorporated by dispersing the cast film with an aqueous solution of the compound. The unencapsulated compound is then removed by centrifugation, chromatography, dialysis or other art-known suitable procedures. The lipid-soluble active ingredient is usually incorporated by dissolving it in the organic solvent with the phospholipid prior to casting the film. If the solubility of the material in the lipid phase is not exceeded or the amount present is not in excess of that which can be bound to the lipid, liposomes prepared by the above method usually contain most of the material bound in the lipid bilayers; separation of the liposomes from unencapsulated material is not required.

A particularly convenient method for preparing liposome formulated forms of hedgehog and ptc therapeutics is the method described in EP-A-253,619, incorporated herein by reference. In this method, single bilayered liposomes containing encapsulated active ingredients are prepared by dissolving the lipid component in an organic medium, injecting the organic solution of the lipid component under pressure into an aqueous component while simultaneously mixing the organic and aqueous components with a high speed homogenizer or mixing means, whereupon the liposomes are formed spontaneously.

The single bilayered liposomes containing the encapsulated hedgehog or ptc therapeutic can be employed directly or they can be employed in a suitable pharmaceutically acceptable carrier for topical administration. The viscosity of the liposomes can be increased by the addition of one or more suitable thickening agents such as, for example xanthan gum, hydroxypropyl cellulose, hydroxypropyl methylcellulose and mixtures thereof. The aqueous component may consist of water alone or it may contain electrolytes, buffered systems and other ingredients, such as, for example, preservatives. Suitable electrolytes which can be employed include metal salts such as alkali metal and alkaline earth metal salts. The preferred metal salts are calcium chloride, sodium chloride and potassium chloride. The concentration of the electrolyte may vary from zero to 260 mM, preferably from 5 mM to 160 mM. The aqueous component is placed in a suitable vessel which can be adapted to effect homogenization by effecting great turbulence during the injection of the organic component. Homogenization of the two components can be accomplished within the vessel, or, alternatively, the aqueous and organic components may be injected separately into a mixing means which is located outside the vessel. In the latter case, the liposomes are formed in the mixing means and then transferred to another vessel for collection purpose.

The organic component consists of a suitable non-toxic, pharmaceutically acceptable solvent such as, for example ethanol, glycerol, propylene glycol and polyethylene glycol, and a suitable phospholipid which is soluble in the solvent. Suitable phospholipids which can be employed include lecithin, phosphatidylcholine, phosphatydylserine, phosphatidylethanol-amine, phosphatidylinositol, lysophosphatidyl-choline and phospha-tidyl glycerol, for example. Other lipophilic additives may be employed in order to selectively modify the characteristics of the liposomes. Examples of such other additives include stearylamine, phosphatidic acid, tocopherol, cholesterol and lanolin extracts.

In addition, other ingredients which can prevent oxidation of the phospholipids may be added to the organic component. Examples of such other ingredients include tocopherol, butylated hydroxyanisole, butylated hydroxytoluene, ascorbyl palmitate and ascorbyl oleate. Preservatives such a benzoic acid, methyl paraben and propyl paraben may also be added.

Apart from the above-described compositions, use may be made of covers, e.g. plasters, bandages, dressings, gauze pads and the like, containing an appropriate amount of a hedgehog or ptc therapeutic. In some cases use may be made of plasters, bandages, dressings, gauze pads and the like which have been impregnated with a topical formulation containing the therapeutic formulation.

EXEMPLIFICATION

The invention now being generally described, it will be more readily understood by reference to the following examples which are included merely for purposes of illustration of certain aspects and embodiments of the present invention, and are not intended to limit the invention.

Muscles are formed by the differentiation of mononucleate proliferative myoblasts into post-mitotic myocytes which subsequently fuse to form multinucleate muscle fibers. In amniotes, muscle fibers differentiate in two waves: the first-formed primary fibers are generally slow, whereas later secondary fibers, which form in close association with primary fibers, are fast (Kelly and Rubinstein, 1980). These two fiber types are not spatially separated and, as they are formed over a considerable time period, the fate of individual cells as they mature has not been followed. In zebrafish, by contrast, somitic muscle fibers form in two temporally-separated waves. The early differentiating cells are formed medially near the notochord and migrate laterally during late somitogenesis to become slow muscle (Devoto et al., 1996; van Raamsdonk et al., 1978). However, most somitic cells differentiate later and become fast muscle.

The differentiation of somites is central to vertebrate mesoderm development. Somites are epithelial balls of mesoderm that arise from a mesenchymal mass of proliferative paraxial tissue in a rostro-caudal order. Once formed, somites rapidly differentiate into a ventral sclerotomal mesenchymal compartment and a dorsal epithelial structure, the dermomyotome. In lower vertebrates, such as fish, in which the sclerotome is small (Morin-Kensicki and Eisen, 1997), the somite mainly gives rise to muscle. In amniotes, the dermomyotome contributes to trunk dermis and to several distinct populations of muscle cells. The dorsomedial lip of the dermomyotome, which is located next to the neural tube, forms the differentiated muscle of the myotome that arises between the sclerotome and dermomyotome.

Signals from adjacent tissues regulate somitic muscle differentiation (for a review see Lassar and Munsterberg (1996)). For instance, Axial structures (neural tube and notochord) are important as their removal leads to cell death and somite regression (Rong et al., 1992; Teillet and Le Douarin, 1983), and they can enhance both myogenesis and chondrogenesis (Kenny-Mobbs and Thorogood, 1987). Notochord can induce myogenesis in some assays of myogenic induction (Buffinger and Stockdale, 1995; Gamel et al., 1995; Stern et al., 1995; Pownall et al., 1996), although ectopically-positioned notochords in chick embryos can induce sclerotome at the expense of

myogenic tissue (Pourquie et al., 1993; Bober et al., 1994; Fan and Tessier-Lavigne, 1994; Goulding et al., 1994). SHH is a signaling molecule expressed in notochord at all times when this tissue can influence muscle differentiation (Echelard et al., 1993; Krauss et al., 1993; Johnson et al., 1994; Roelink et al., 1994). SHH can substitute for notochord in various assays of both sclerotome and muscle induction (Fan et al., 1995; Munsterberg et al., 1995), and to induce ectopic muscle markers in vivo (Johnson et al., 1994; Concordet et al., 1996; Weinberg et al., 1996; Hammerschmidt et al., 1996). Moreover, mice homozygous for a targeted deletion of the *shh* gene have deficits in sclerotome and myotome precursor cell markers (Chiang et al., 1996). These data suggest that SHH may mediate notochord-dependent signals that induce myogenesis. However, two lines of evidence argue against this simple view. First, both the MyoD and Myf-5 muscle specific transcription factors are still expressed in $shh^{-/-}$ mouse somites, although *Myf-5* mRNA is reduced (Chiang et al., 1996). As Myf-5 and MyoD are myoblast markers in amniotes, this suggests that the myogenic program can be initiated in the absence of SHH. Second, ablation of all axial structures has little effect on limb and body wall muscle development although somitic myogenesis is reduced, partly due to regression of the somite (Rong et al., 1992). Thus, although notochord-derived SHH is a strong candidate for a regulator of myotomal muscle formation, its precise role in myogenesis has previously remained enigmatic.

A confounding factor in understanding myotomal muscle induction is the heterogeneity of myogenic cell populations within the somite (reviewed in Cossu et al. (1996b)). The phenotypes of mice with null mutations in members of the MyoD family of myogenic regulatory transcription factors (MRFs) suggest that several distinct populations of myogenic cells exist in different parts of the developing murine dermomyotome (Rudnicki et al., 1993; Tajbakhsh et al., 1997) and these populations appear to differ in their sensitivity to loss of SHH (Chiang et al., 1996). In addition to notochord, neural tube also contains inductive signals that can support somitic myogenesis (Buffinger and Stockdale, 1995; Goulding et al., 1994; Rong et al., 1992; Stern and Hauschka, 1995; Teillet and Le Douarin, 1983), and dorsal neural tube can induce myogenesis, an effect that can be mimicked by some Wnt proteins (Gamel et al., 1995; Munsterberg et al., 1995; Stern et al., 1995). Moreover, inhibitory signals from lateral plate mesoderm and surface ectoderm have been suggested to influence myogenesis (Fan and Tessier-Lavigne, 1994; Pourquie et al., 1996). Thus, although several distinct signals and muscle cell populations exist, what signals induce each cell population in vivo is unclear.

In the zebrafish, the somite gives rise mainly to muscle, which is probably the primary fate of paraxial mesoderm during early chordate evolution (Holland et al., 1995). Even in this simple system, however, three muscle cell populations can be resolved. Adaxial cells form next to the notochord, eventually giving rise to slow muscle (Devoto et al., 1996; van Raamsdonk et al., 1978), and are the first cells in the embryo to express the muscle transcription factors *myoD* and *mef2D, A* and *C* (Ticho et al., 1996; Weinberg et al., 1996). A specialized subpopulation of adaxial cells, the muscle pioneers, form at the dorsoventral midline of each somite (Felsenfeld et al., 1991), express engrailed proteins (Hatta et al., 1991), and appear to be induced by two sequential signals from outside the somite (Currie and Ingham, 1996). The majority of the somite forms the third muscle cell population that both expresses *myoD* and differentiates later (Devoto et al., 1996; Weinberg et al., 1996).

Methods

Zebrafish lines and maintenance

Wild-type and heterozygote mutant breeding fish were maintained at 28.5 C on a 14-hour/10-hour light cycle. We obtained *floating head*^{n1} *(flh)* from the University of Newcastle upon Tyne, *notail*^{b160} (ntl) from the University of Oregon, and *boxozok*^{i2} *(boz)* was isolated in the Ingham laboratory (P.D.C., T. Schilling, G. Bergemann and P.W.I., unpublished data). boz^{i2} fish exhibit a variable phenotype with defects ranging from reduced notochords to a severe lack of axial mesoderm at all rostro-caudal levels. Of 142 progeny of a hetrozygous cross examined at F_2-F_4, 20 (13%) showed complete absence of eyes and notochord. A further 25 (17%) showed a partial phenotype with variable eyes and the anterior half of the notochord missing. These, and a number of other aspects of the phenotype, are strongly reminiscent of bozozok mutant fish (Solnica-Krezel et al., 1996). Complementation analysis by a cross of heterozygous boz^{i2} with boz^{m168} has shown reduced eyes and notochord in three out of 30 progeny. We therefore tentatively conclude that these genes are allelic. However, due to the incomplete penetrance of *boz*, definitive demonstration of allelism awaits the mapping of the mutation. Embryos were collected by natural spawning and staged by anatomical markers according to Westerfield (1995). *Prim-5* embryos are referred to as 24 hour.

RNA injection

RNA injections were performed as described (Currie and Ingham, 1996).

Immunohistochemistry

The Slow and fast MyHC antigens are destroyed by aldehyde fixatives, so embryos were fixed by incubating for 5 minutes each in graded methanols, rehydrated in 0.1% Tween-20, serially cryosectioned and stained. However, preservation of younger embryos was better after staining in whole-mount, followed by post-fixation in 4% paraformaldehyde for 4 hours at 4 C prior to cryosectioning. Primary monoclonal antibody supernatants of A4.1025 (Dan-Goor et al., 1990) and BA-D5 (Schiaffino et al., 1989) were diluted 1:10. EB165 monoclonal ascites was used at 1:5000 (Gardahaut et al., 1992). First antibodies were detected with biotin-conjugated horse-derived anti-mouse IgG (Vector), Vectastain ABC Elite Peroxidase kit (Vector) and visualized using 0.5 mg/ml diaminobenzidine with (black stains) or without (brown stains) 0.03% $CoCl_2$ enhancement. Cryosections for dual immunofluorescence had IgG first antibodies detected with Cappell goat anti-mouse IgG (y-specific) Texas red. After a mouse IgG block, biotinylated BA-D5, prepared using Pierce NHS-Biotin reagent, was detected with Dako streptavidin-FITC. Sections were mounted in 150 mg/ml polyvinyl alcohol 30% glycerol PBS with DABCO antifade, and photographed by confocal microscopy.

Western blots

Embryos were dechorinated, deyolked and homogenized manually on ice for 10 minutes in 63 mM Tris-HCl pH 6.8. 10% glycerol, 5% B-mercaptoethanol, 3.5% SDS, 0.2 mM PMSF, 0.5 M aprotinin, 0.5 M leupeptin. Samples were microfuged for 5 minutes at 4 C, 0.01% bromophenol blue added to the supernatant, the equivalent of 10 embryos run on each lane of a 7.5% acrylamide denaturing gel at 200mV for 30 mins and electroblotted onto nitrocellulose (Amersham). Purified bovine cardiac myosin was a kind gift of Dr. John Sleep. Nitrocellulose strips were blocked in 5% milk powder PBS/Az overnight, washed, and incubated with A4.0125 (1:10), BA-D5 (1:10), F1.652 (1:10 (Webster, et al. 1988)) or EB 165 (1:250) for 2h at RT. After washing, primary antibody was detected with horseradish peroxidase-conjugated sheep anti-mouse IgGF(ab)2 and an ECL kit (Amersham).

Results

Early zebrafish embryos have distinct fast and slow muscle cell populations

To examine the patterning of muscle in zebrafish somites we screened a series of anti-MyHC monoclonal antibodies for reactivity with 1-2 day zebrafish muscle tissue.

Two antibodies detected all differentiated skeletal and heart muscle, whereas three antibodies detected specific subpopulations of cells within the somites of 24 hour (prim-5) embryos. BA-D5, an antibody that specifically detects slow MyHC in muscle fibers of all ages of mammals and chicks examined (Schiaffino et al. (1989) and our unpublished observations), detects a single layer of cells in the superficial region of 24 hour zebrafish somites at all antero-posterior positions within the body axis. In contrast, EB 165, an antibody that detects fast fibers in embryonic and adult chicken muscle (Gardahaut et al., 1992), detects an adjacent non-overlapping population of medial somitic muscle fibers in the one day zebrafish embryo. A third monoclonal antibody, A4.1025, which reacts with a conserved epitope near the ATP-binding site of all striated muscle MyHC isoforms examined in a wide variety of species (Dan-Goor et al., 1990), detects both the BA-D5+ and EB165+ populations of cells. All three antibodies reacted with muscle fibers in a striated pattern typical of sarcomeric myosin and Western analysis of 24 hour zebrafish extracts separated by SDS PAGE demonstrated that all three antibodies detect protein bands at or just under M_r 200 000, the size of MyHC osoforms. Thus, these anti-MyHC antibodies distinguish slow and fast differentiated muscle cells in the zebrafish embryo.

Slow muscle differentiates before fast in zebrafish embryos

We determined the timing and location of slow and fast muscle differentiation throughout zebrafish somite development. New somites separate from the presomitic paraxial mesoderm in an anterior to posterior order about every half hour between 10.5 and 26 hours of development at 28 C (Westerfield, 1995. We observed that MyHC+ adaxial cells appeared on each side of the notochord in an anterior to posterior order in each somite as it formed. Most, if not all, adaxial MyHC+ cells also express slow MyHC. Fast MyHC was undetectable in 15 somite embryos. Thus, the first population of muscle cells to differentiate in the zebrafish somite are the adaxial cells, and these cells express slow, but not fast, characteristics from their inception.

A recent study by Devoto et al. (1996) has elegantly shown that the differentiated adaxial cells of somites 16-20 in the gut extension region of zebrafish embryos migrate laterally through the somite 3-4 hours post-somitogenesis. Consistent with the results of Devoto et al., we find that prior to the 20 somite stage adaxial slow MyHC+ cells in all somites remain medial, but that the adaxial slow MyHC+ cells of older somites appear to spread dorsally around the sides of the neural tube and ventrally past the hypochord to form a single layer of medial cells. At about the 20 somite stage the adaxial slow MyHC+ cells of the most anterior somites appear to migrate laterally, through the

undifferentiated somitic mesoderm. Although it is possible that this apparent migration represents a wave of fiber type conversion, we think this unlikely from the earlier findings of Devoto et al. (1996) that early adaxial cells migrate and form slow muscle. The wave of migration sweeps rapidly along the embryo from anterior to posterior so that by the 21 somite stage slow muscle cells of the anterior somites are located at the lateral edge of the somite under the epidermis, whereas slow muscle cells of mid-body somites are found in the center of the somite and the most posterior slow muscle cells are still in the adaxial position. During the lateral migration of slow muscle cells the differentiation of fast muscle cells commences. No differentiated fast muscle was observed lateral to migrating slow muscle cells. However, strikingly, fast muscle cells are detected medial to the slow muscle cells immediately after the migratory period in each somite. By the 26 somite stage, all slow muscle cells in somites 1-21 have migrated laterally. At this stage fast muscle fills the medial bulk of the somite. Thus, the differentiation of a distinct class of fast muscle cells rapidly succeeds the migration of the slow muscle cells past undifferentiated somitic cells.

Notochord defects correlate with lack of slow muscle differentiation

The formation of slow muscle next to notochord suggests that a notochord derived signal induce slow muscle cells. To test this hypothesis we examined two mutant zebrafish strains that are defective in distinct stages of notochord development. Severely affected *bozozok* (boz^{i2}) fish do not have visible notochord, lack the notochord and floorplate marker SHH, and completely lack differentiated slow muscle. At 24 hours of development, when in wild type embryos adaxial cells have differentiated, migrated and express slow MyHC in all somites, no slow MyHC is detected in trunk or tail regions of *boz* embryos that lack notochord (6 of 6 embryos sectioned), although unaffected sibling embryos appear wild type (data not shown). Mutation of the *boz* gene does not prevent muscle differentiation *per se* because a single fused somite of differentiated muscle is present beneath the neural tube, and this expresses fast MyHC. Normal *boz* function is required for formation of slow muscle, rather than maintenance, as both severely and more mildly affected embryos from a *boz* heterozygote cross failed to express detectable slow MyHC at the 15 somite stage, whereas morphologically normal siblings showed normal slow MyHC expression (data not shown). Thus, the absence of notochord in the *boz* mutant is accompanied by the specific loss of slow muscle.

Although *boz* function is required for notochord formation, it is possible that the wild type gene might also be required in paraxial mesoderm to permit differentiation of

slow muscle cells. We therefore examined *no tail* (*ntl*) mutant embryos in which midline mesodermal cells are present but fail to differentiate into mature notochord cells. Previous studies have shown that *ntl* embryos also lack muscle pioneer cells, a sub-population of the adaxial cells (Halpern et al., 1993). *ntl* is the zebrafish homologue of the Brachyury transcription factor and is expressed in notochord but not in adaxial cells at the time of their differentiation and hence muscle defects are unlikely to be due to a cell-autonomous action of *ntl* in paraxial mesoderm (Schulte-Merker et al., 1994). In *ntl b160* embryos, notochord precursors are present in anterior regions but absent posteriorly in the region beyond the yolk tube which is severely truncated (Halpern et al., 1993; Odenthal et al., 1996). We examined *ntl*^{b160} fish for slow MyHC expression anticipating that the loss of notochordal maturation might prevent slow muscle formation. Despite the absence of muscle pioneer cells at the dorsoventral midline, slow and fast muscle in anterior regions of *ntl* embryos appeared normal. Thus, mature notochord is not required for slow muscle differentiation. However, more posterior regions of *ntl* embryos, in which axial mesoderm defects are more severe (Halpern et al., 1993), showed reduced slow muscle formation and aberrant positioning. In rare embryos (1/16 serially sectioned) a complete absence of slow muscle was observed in the most posterior somite at 24 hours of development, even though extensive differentiated fast muscle was present. The remaining *ntl* embryos (15/16) showed regional slow deficits. Thus, *ntl* mutant fish demonstrate that although mature notochord is not necessary for slow muscle formation, severe defects in notochord establishment in the tail correlate with loss of slow muscle differentiation.

Sonic hedgehog induces ectopic slow muscle differentiation

Examination of the muscle phenotype of *boz* and *ntl* mutant embryos suggested that notochord-derived signals may determine the slow muscle fate, reminiscent of the induction of floorplate and motoneurons by notochord-derived SHH protein (Ericson et al., 1996) and of muscle pioneer cells by notochord-derived hedgehogs (Currie and Ingham, 1996). Consistent with this *shh* mRNA is absent from those regions of both *boz* and *ntl* embryos that show defects in slow muscle formation. To test the possibility that SHH might be a notochord-derived inducer of the slow muscle fate, we injected *shh* mRNA into two or four cell zebrafish embryos to create animals chimeric for *shh* over-expressing cells. Such injections lead to an easily detectable reduced retina phenotype (Krauss et al., 1993). In animals affected for retinal development, we observed an induction of slow MyHC expression across the entire width of the somite in each of twelve serially-sectioned 24 h embryos. Strikingly, this expansion occurs at the expense

of differentiated fast muscle. In those animals in which mosaic segregation of *shh* mRNA causes partial slow muscle induction, residual fast muscle is observed in regions not expressing slow MyHC (data not shown). When the same experiment was repeated using an equivalent amount of *echidna hedgehog* (*ehh*) mRNA no defect was detected in any part of ten embryos serially sectioned (data not shown). Thus, SHH is a notochord-derived signal capable of inducing slow muscle at the expense of fast.

The wholesale conversion of large areas of somite to slow muscle by SHH has two possible explanations. SHH could induce somitic cells to differentiate as slow muscle prematurely. Alternatively, SHH might not affect the decision of when to differentiate, but simply determine what type of muscle is formed. To address this issue we examined the effect of ectopic SHH on earlier stage zebrafish embryos. In 15 somite zebrafish embryos, SHH induces a wide region of ectopic lateral differentiated muscle within the somite (46/53 unselected injected embryos). Ectopic slow muscle differentiation occurred without premature induction of fast muscle tissue. The premature differentiation of lateral muscle tissue suggested that SHH might induce presomitic mesoderm to differentiate early. However, premature slow muscle differentiation before the normal time of adaxial cell differentiation was not observed in either presomitic mesoderm of any of 36 embryos examined at the 15 somite stage or in any region of 22 embryos at tailbud stage (data not shown). Therefore, the earliest time somitic cells are competent to become slow muscle in response to SHH is when adaxial cells normally differentiate, but at this stage cells in all regions of the somite become competent.

The ability of SHH to induce slow muscle is consistent with the lack of slow muscle in regions of zebrafish mutants that lack midline *shh* expression. This correlation strongly suggests that the reason for the lack of slow muscle in the *boz*^{i2} and the tail of *ntl*^{b160} mutants is lack of notochord-derived SHH (Concordet et al., 1996). However, the *boz* gene has not been cloned, so its expression is unknown, and the *ntl* gene is expressed transiently in presomitic mesoderm, as well as in notochord (Odenthal et al., 1996). This raises the possibility that the lack of slow muscle in these mutants reflects a need for a cell autonomous action of the respective genes in paraxial mesoderm. To address this issue, we over-expressed *shh* in embryos from *boz*^{i2} mutant crosses and examined the resultant animals for slow MyHC expression. Five out of six severely affected *boz* mutants injected with *shh* mRNA showed induction of slow MyHC, and suppression of fast MyHC. Thus, the *boz*^{i2} mutation does not affect the ability of somitic tissue to respond to SHH and form slow muscle. Moreover, even in the complete absence of notochord SHH is sufficient for the formation of slow muscle.

A limited source of SHH is sufficient to rescue slow muscle

One limitation of over-expression of SHH by mRNA injection is that ectopic SHH is expressed in regions of the animal never normally exposed to SHH and at above normal physiological levels. This might perturb signals necessary for muscle development from other embryonic tissues. To examine slow muscle differentiation in response to localized lower levels of shh expression we took advantage of the *floating head* (*flh*) mutation. Animals homozygous for *flh* are defective in notochord (Halpern et al., 1995; Talbot et al., 1995) and, like *ntl* embryos, lack notochords and muscle pioneers. However, unlike *ntl*, *flh* exhibit transdifferentiation of notochord tissue into muscle (Halpern et al., 1995). We examined *flh* embryos for muscle differentiation and found that it occurs in an altered location. Cells in the embryonic midline, not those in the adaxial position are the first to differentiate in *flh* embryos. This differentiation is immediately beneath the presumptive floorplate that expresses SHH sporadically. Despite the unusual location of these muscle cells, they express slow but not fast MyHC, spread dorsally around the neural tube and ventrally in the midline and appear able to undergo lateral migration to take up a normal position beneath the ectoderm by 24 hours of development. Thus, in *flh* embryos, apparently normal slow muscle cells differentiate beneath the residual floorplate: the sole remaining location where somitic mesoderm abuts *shh*-expressing tissue.

The discontinuous location of *shh*-expressing cells in the floorplate of *flh* mutants allows an examination of the relationship between the location of SHH and slow muscle differentiation. At the posterior limit of MyHC-containing cells in *flh* we found no correlation between the location of remaining floorplate *shh* expression and medial slow myoblast differentiation: muscle differentiates both immediately beneath and between islands of *shh*-expression. This data suggested that terminal differentiation of slow muscle is not directly induced by SHH. Further evidence that muscle differentiation per se is not induced by SHH came from examining the up-regulation of *patched I (ptcl)* mRNA in *flh* mutant embryos. *Ptcl.*a zebrafish homologue of Drosophilia *patched.* Is a SHH receptor (Stone, et al. 1996), and is up-regulated adjacent to residual *shh* expression in *flh* embryos both at somitic and presomitic antero-posterior positions (Concordet et al., 1996). Therefore, mesodermal cells are first exposed to SHH long before muscle differentiation commences. Moreover, event the most recently differentiated muscle cells in *flh* embryos frequently do not express high levels of *ptcl* mRNA, despite adjacent mesoderm expressing *ptcl* abundantly. Thus, although SHH induces *ptcl* locally along the entire length of the *flh* embryo, there is a delay after SHH exposure before the appearance

of differentiated slow muscle cells. In addition, by the time slow muscle differentiates, any spatial correlation between *shh*-expression and myogenic cells has been lost. Taken together, these data suggest that SHH may initiate slow myoblast formation, but that continued exposure is not required to trigger the terminal differentiation of slow muscle fibers.

Discussion

Sonic hedgehog and slow muscle induction

Several lines of evidence show that adaxial slow muscle formation in zebrafish embryos is controlled by SHH. First, slow muscle differentiation occurs next to the notochord, which expresses *shh*. This observation confirms and extends the results of Devoto et al. (1996), who observed that adaxial cells give rise to slow muscle markers after they migrate laterally through the somite. Our data show that adaxial cells are already determined to form slow muscle as soon as they differentiate into skeletal muscle myosin-expressing myocytes. Second, a lack of slow muscle correlates with a lack of *shh* expression in mutuant fish. Third, ectopic expression of *shh* can induce conversion of most, if not all, somitic cells to slow muscle, at the expense of fast muscle. Fourth, even in the absence of notochord, injection of *shh* mRNA can rescue formation of slow muscle. Fifth, in the absence of the normal SHH signal from notochord, a localized source of SHH correlates with induction of ectopic slow muscle cells, which then migrate in a similar fashion to slow muscle cells in wild type embryos. These data are supported by the wild type expression of a SHH receptor, *ptcl.* That is upregulated in adaxial cells within presomitic mesoderm indicating that these cells are responding to hedgehog signaling (Concordet et al., 1996). Taken together, these findings made a strong case for notochord-derived SHH being the normal inducer of the differentiated slow adaxial muscle cell fate in the zebrafish.

Sonic hedgehog induces adaxial slow myoblasts

How might SHH induce the slow muscle fate in zebrafish? Muscle is formed in two steps: mesodermal commitment to the proliferative myoblast, followed by terminal differentiation into the post-mitotic muscle fiber. Several lines of evidence suggest that SHH is responsible for induction of slow muscle precursor cells, rather than the terminal differentiation of slow muscle *per se.*

First the muscle-specific transcription factor *myoD* is initially detectable in adaxial precursors located adjacent to *shh*-expressing cells within the embryonic shield several hours before their terminal differentiation at around the time of somitogenesis (Weinberg et al. 1996). This expression of *myaD* prior to terminal differentiation is also detected in adaxial cells at later stages when posterior somites arise from the tail bud. Second, zebrafish mutants like *boz* and *ntl* that lack slow muscle, also lack the early adaxial *myoD* expression, and this correlates with a lack of axial SHH (Concordet et al., 1996; Odenthal et al., 1996; Weinberg, et al. 1996; Schier et al., 1997). Third, SHH signaling can induce premature *myoD* in lateral presomitic cells (Concordet et al., 1996; Hammerschmidt et al., 1996; Weinberg et al., 1996). We show that these ectopic *myoD*-expressing cells in lateral somite have other features, such as the direction and timing of their differentiation and sensitivity to additional hedgehog signals (Currie and Ingham, 1996), suggesting that the terminal differentiation of these cells into slow muscle is pre-figured at the myoblast level. Fourth, our examination of the *flh* mutuant suggests that adaxial myoblasts differentiate into slow muscle fibers independent of their proximity to residual *shh*-expressing floor plate cells, and independent of their exposure to SHH during the period of terminal differentiation, as assayed by *ptcl* expression (Concordet et al., 1996; Marigo and Tabin, 1996). Thus, at early stages MyoD may mark cells that, while not yet differentiated, have become committed to a slow myoblast lineage.

Previous data has suggested that the combined action of notochord-derived Sonic and Echidna hedgehogs induces zebrafish muscle pioneer subset of the adaxial slow muscle cells *Currie and Ingham, 1996). The data in the present paper demonstrate that EHH is not required for production of the non-pioneer adaxial slow muscle cells. EHH is not expressed in notochord of $ntl^{b160'}$ (Currie and Ingham, 1996), yet non-pioneer slow muscle cells form and migrate normally in the anterior of ntl^{b160} embryos where SHH alone is expressed. Similarly, in *flh* mutants, which lack notochord and *ehh* expression (Currie and Ingham, 1996), apparently normal non-pioneer adaxial cells are formed ectopically close to residual floorplate SHH. Moreover, EHH does not appear able to substitute for SHH in the induction of non-pioneer adaxial cells as injection of *ehh* mRNA into wild type embryos did not induce ectopic slow MyHC. These data support the hypothesis that in vivo SHH and EHH serve distinct roles.

The finding that SHH induces slow myoblasts suggests a new view of the steps of muscle differentiation that contrasts with the traditional model in which somitic cells first become myoblasts and only subsequently specialize into one particular myoblast sub-class. We that the decision whether to form one type of muscle or another is made concurrently with myoblast commitment to the muscle lineage. This scheme concurs

with conclusions from analysis of the *you-* type zebrafish mutants (van Eeden et al., 1996). Such a veiw also fits well with studies in *Drosophila* demonstrating that distinct extracellular signals serve to commit each founder myoblast to a particular muscle type (Baylies et al., 1995). However, it is possible that presomitic cells could be committed to myogenesis prior to *myoD* expression. Although MyoD is the first MRF to be expressed in birds, Myf-5 is the earliest MRF to appear at high levels in mammalian somites (Ott et al., 1991), and Pax-3 can induce myogenesis (Tajbakhsh et al., 1997; Maroto et al., 1997). Furthermore, whether all myoblasts are committed to form particular types of muscle from their inception is unclear. Whatever the case, our data shows that SHH induces adaxial myoblasts that adopt a slow muscle fate.

Zebrafish fast muscle formation

SHH is not necessary for fast muscle formation. *Boz* fish that lack SHH produce abundant fast muscle throughout the somite. Moreover, the normal *myoD*-expressing myoblasts stripes across the posterior border of the somite form at the normal time just prior to somitogenesis in embryos that lack *shh* expression (Odenthal et al., 1996). Thus, in zebrafish, MyoD may mark commitment to a myoblast fate irrespective of the type of myoblast formed. We find that cells that are initially lateral within the somite differentiate into fast muscle: they may be committed to formation of fast muscle from the inception of *myoD* expression.

Timing of myoblast differentiation

We show that extopic SHH can induce premature muscle differentiation in the lateral somite. However, premature differentiation was slow, rather than fast, and was only observed at the normal time of slow adaxial cell differentiation. Therefore, no somite cells are competent to differentiate in response to SHH until around the time of somitogenesis, even though *myoD* is expressed earlier. This may explain why slow muscle does not appear earlier in development even though *shh* is expressed in the developing notochord from gastrulation onwards (Krauss et al., 1993), and is presumably secreted because *ptcl*, a marker of SHH exposure, is highly expressed in adjacent presomitic cells (Concordet et al, 1996). Two alternative models could explain the delay between MyoD and slow myosin expression. In one model the delay is due to an intrinsically timed maturation of the somitic cells. Although cell division is not extensive in zebrafish somites (Kimmel and Warga, 1987), SHH might induce myoblasts committed to division followed by differentiation as it can be a somitic mitogen (Fan et

al., 1995). Mammalian myoblasts show such behavior in vitro (Quinn et al., 1985), which is reminiscent of the induction of division followed by terminal differentiation in Drosophila lamina ganglia neurons in response to retinal neuron-derived hedgehog (Huang and Kunes, 1996). Alternatively, in the second model, an extracellular signal(s) may control terminal differentiation. In amniotes, other signals can cooperate with SHH to regulate myogenesis (Munsterberg et al., 1995; Pourquie et al., 1996; Stern et al., 1995), and a variety of growth factors repress myoblast differentiation in culture. Ventral axial structures are unlikely sources of such signals as SHH is sufficient to induce slow MyHC in *boz* embryos that fail to form notochord or floor plate. Regardless of the mechanism by which the timing of terminal adaxial slow muscle differentiation is controlled, our data show that similar mechanisms can operate in the lateral somite to control ectopic slow muscle differentiation in response to SHH.

Evolutionary conservation of muscle patterning

We found that in fish embryos the first skeletal muscle fibers to form are slow from the time of their inception. Later, a second wave of fibers, which ultimately constitute the majority of all fibers, differentiate as fast muscle. In amniote limbs muscle fibers also form in two waves, an early primary population that express slow (and embryonic) myosin and later secondary cells that, forming in close association with primary fibers, express fast (and embryonic) myosin from their inception (Kelly and Rubinstein, 1980: Vivarelli et al., 1988: Cho et al., 1994). We suspect that zebrafish embryos may also express and embryonic myosin in both slow and fast fibers as the immunoreaction with the our all-myosin antibody was stronger than with the specific slow and fast antibodies. Moreover, both adaxial and non-adaxial cells react from their inception with an antibody that detects embryonic myosin (Devoto et al., 1996). These analogies suggest that adaxial and non-adaxial somitic muscle cells in the zebrafish may be evolutionary homologues of amniote primary and secondary muscle fiber generations. Amniote secondary fibers form overlying the neuromuscular junctions of primary fibers and it has been suggested that signals from the forming neuromuscular junction region may be required to initiate secondary fiber formation (Duxson et al., 1989). This is not the case in the zebrafish as absence of differentiated slow primary fibers does not prevent differentiation of fast muscle despite the striking correlation between the lateral migration of slow fibers and the differentiation of fast fibers. The converse relationship, that fast fiber differentiation might cause slow fiber migration, remains a possibility. Nevertheless, the close similarities between fish and amniote fiber generation suggest that the common ancestor had two steps of muscle patterning: early fibers being slow and

later fast. There are further analogies between amniote and fish myogenesis. Amniote primary fibers are of several distinct fiber types that prefigure later muscle characteristics (Crow and Stockdale, 1986), even though all express some form of slow MyHC (Kelly and Rubinstein, 1980; Vivarelli et al., 1988; Page et al., 1992; Hughes et al., 1993). Slow adaxial cells in the zebrafish are also composed of two sub-populations, the muscle pioneer cells which express engrailed, and the non-pioneer adaxials. Engrailed proteins also mark a sub-population of muscle cells in the jaw muscle of the zebrafish (Hatta et al., 1990). The data reported i the present paper, together with the previous findings that a second notochord-derived signal (Halpern et al., 1993), provided by EHH (Currie and Ingham, 1996), is responsible for regulating the formation of muscle pioneer cells, suggest that hedgehog signalling molecules may regulate the diversity of muscle fiber types formed in the early fish embryo. *Banded hedgehog* is also expressed in particular regions of the Xenopus somite (Ekker et al., 1995). Whether similar signals control muscle patterning in amniotes remains to be determined.

Hedgehogs and vertebrate myogenesis

Secretion of SHH from notochord has been shown to induce floorplate markers in anterior (although not posterior) zebrafish CNS and both floorplate and motoneurons in amniote neural tube (Echelard et al., 1993; Krauss et al.. 1993; Roelink et al.. 1994; Ericson et al., 1996). The role of SHH in somite patterning has been less clear. In amniotes, in which much of th esomite becomes sclerotome, either octopic notochord or *shh*-expressing cells can induce extra sclerotome at the expense of dermomyotome markers (Fan and Tessier-Lavigne, 1994). Conversely, *shh*$^{-1-}$ mice have deficits in sclerotomal derivatives (Chiang et al., 1996). On the other hand, in both chick and zebrafish, over-expression of *shh* induces ectopic *myoD* expression, suggesting a myogenic action (Johnson et al., 1994; Concordet et al., 1996; Weinbert et al., 1996). Moreover, *shh*$^{-1-}$ show defects in medial muscle formation (Chiang et al., 1996) and notochord can induce avian myogenesis (Pownall et al., 1996). So SHH May regulate formation of both ventral and more dorsal somitic tissues. Action of SHH at distinct concentrations or times (Ericson et al., 1996), or in collaboration with other factors (Munsterberg et al., 1995; Stern et al., 1995; Pourquie et al., 1996) could determine the outcome of SHH signaling.

Induction of distinct myoblast types and the subsequent control of their terminal differentiation account for the numerous signals capable of influencing somite myogenesis. If equivalents of SHH-dependent adaxial cells exist in amniotes, we would

expect that particular muscle markers are not uniformly distributed between distinct muscle cell types in the developing dermomyotome. In amniotes MRFs are the earliest known definitive myogenic markers. Expression of at least one MRF is obligatory for myogenesis in mice (Rudnicki et al., 1993). MRFs are expressed at low levels in presomitic mesoderm which has the capacity to form muscle in dissociated cell culture (George-Weinstein et al., 1994, Lin-Jones and Hauschka, 1996). However, two myoblast populations arise with distinct temporal and spatial patterns within the dermomyotome: the first initially expresses *myf-5* in medial, and the second *myoD* in lateral regions (Cossu et al., 1996a; Tajbakhsh et al., 1997; Maroto et al., 1997). That *shh*$^{-/-}$ mice have reduced expression of medial *myf-5* but no detectable change in lateral *myoD* expression (Chiang et al., 1996) suggests a role for SHH in induction of the medial population. Inhibitory signals, such as BMP4 (Fan and Tessier-Lavigne, 1994; Cossu et al., 1996b; Pourquie et al., 1996), may function in vivo to suppress overt myogenic phenotypes in the lateral compartment that generates limb and body wall muscle and may have no homologous process in most zebrafish somites. So generation of further diversity within the dorsomedial myogenic compartment could be a role of SHH in amniote myogenesis. Distinct populations of slow and fast fibers may be present in amniote myogenesis. Distinct populations of slow and fast fibers may be present in amniote myotome (Dhoot, 1994). In this paper, we have shown that in zebrafish SHH regulates formation of myotomal slow muscle. Much slow muscle in amniote limbs is located near developing bone that expresses *indian hedgehog* (Vortkamp et al., 1996; Bitgood and McMahone, 1995). Moreover, motoneurons, which strongly influence muscle development, can express *shh* (Bitgood and McMahone 1995; Stone et al., 1996), raising the possibility that diverse hedgehog proteins may regulate muscle fiber diversification.

References Cited in Background and Examples:

Baylies, M.K., A. Martinez-Arias and M. Bate. 1995. *wingless* is required for the formation of a subset of muscle founder cells during *Drosophila* embryogenesis. *Development*. *121*: 3829-3837.

Bitgood, M.J. and A.P. McMahon. 1995. ~~Hedgehog and Bmp genes are coexpressed at many~~ diverse sites of cell- cell interaction in the mouse embryo. *Dev.Biol.* *172*: 126-138.

Bober, E., B. Brand-Saberi, C. Ebensperger, J. Wilting, R. Balling, B.M. Paterson, H.-H. Arnold and B. Christ. 1994. Initial steps of myogenesis in somites are independant of influence from axial structures. *Development*. *120*: 3073-3082.

Buffinger, N. and F.E. Stockdale. 1995. Myogenic specification of somites is mediated by diffusible factors. *Dev.Biol.* *169*: 96-108.

Chiang, C., Y. Litingtung, E. Lee, K.E. Young, J.L. Corden, H. Westphal and P.A. Beachy. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature*. *383*: 407-413.

Cho, M., S.M. Hughes, I. Karsch-Mizrachi, M. Travis, L.A. Leinwand and H.M. Blau. 1994. Fast myosin heavy chains expressed in secondary mammalian muscle fibres at the time of their inception. *J.Cell Sci.* *107*: 2361-2371.

Concordet, J.-P., K.E. Lewis, J.W. Moore, L.V. Goodrich, R.L. Johnson, M.P. Scott and P.W. Ingham. 1996. Spatial regulation of a zebrafish *patched* homologue reflects the roles of *sonic hedgehog* and protein kinase A in neural tube and somite patterning. *Development*. *122*: 2835-2846.

Cossu, G., R. Kelly, S. Tajbakhsh, S. Di Donna, E. Vivarelli and M. Buckingham. 1996a. Activation of different myogenic pathways: myf-5 is induced by the neural tube and MyoD by the dorsal ectoderm in mouse paraxial mesoderm. *Development*. *122*: 429-437.

Cossu, G., S. Tajbakhsh and M. Buckingham. 1996b. How is myogenesis initiated in the embryo? *Trends Genet.* *12*: 218-222.

Crow, M.T. and F.E. Stockdale. 1986. Myosin expression and specialization among the earliest muscle fibres of the developing avian limb. *Dev.Biol.* *113*: 238-254.

Currie, P.D. and P.W. Ingham. 1996. Induction of a specific muscle cell type by a hedgehog-like protein in zebrafish. *Nature*. *382*: 452-455.

Dan-Goor, M., L. Silberstein, M. Kessel and A. Muhlrad. 1990. Localization of epitopes and functional effects of two novel monoclonal antibodies against skeletal muscle myosin. *J. Muscle Res. Cell Motil.* *11*: 216-226.

Devoto, S.H., E. Melancon, J.S. Eisen and M. Westerfield. 1996. Identification of separate slow and fast muscle precursor cells in vivo, prior to somite formation. *Development*. *122*: 3371-3380.

Dhoot, G.K. 1994. Mammalian myoblasts become fast or slow myocytes within the somitic myotome. *J.Muscle Res.Cell Motil.* *15*: 617-622.

DiMario, J.X., S.E. Fernyak and F.E. Stockdale. 1993. Myoblasts transferred to the limbs of embryos are committed to specific fibre fates. *Nature*. *362*: 165-167.

Duxson, M.J., Y. Usson and A.J. Harris. 1989. The origin of secondary myotubes in mammalian skeletal muscles: ultrastructural studies. *Development*. *107*: 743-750.

Echelard, Y., D.J. Epstein, B. St.-Jaques, L. Shen, J. Mohler, J.A. McMahon and A.P. McMahon. 1993. Sonic hedgehog, a member of a family of putative signalling molecules implicated in regulation of CNS and limb polarity. *Cell*. *75*: 1417-1430.

Ekker, S.C., L.L. McGrew, C.J. Lai, J.J. Lee, D.P. Von Kessler, R.T. Moon and P.A. Beachy. 1995. Distinct expression and shared activities of members of the hedgehog gene family of Xenopus laevis. *Development*. *121*: 2337-2347.

Ericson, J., S. Morton, A. Kawakami, H. Roelink and T.M. Jessell. 1996. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell*. *87*: 661-673.

Fan, C.-M., J.A. Porter. C. Chiang. D.T. Chang, P.A. Beachy and M. Tessier-Lavigne. 1995. Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cAMP signalling pathway. *Cell. 81*: 457-465.

Fan, C.-M. and M. Tessier-Lavigne. 1994. Patterning of mammalian somites by surface ectoderm and notochord: evidence for sclerotome induction by a hedgehog homolog. *Cell.* ~~79: 1175-1186.~~

Felsenfeld, A.L., M. Curry and C.B. Kimmel. 1991. The *fub-1* mutation blocks initial myofibril formation in zebrafish muscle pioneer cells. *Dev.Biol. 148*: 23-30.

Gamel, A.J., B. Brand-Saberi and B. Christ. 1995. Halves of epithelial somites and segmental plate show distinct muscle differentiation behavior in vitro compared to entire somites and segmental plate. *Dev. Biol 172*: 625-639.

Gardahaut, M.F., J. Fontaine-Perus, T. Rouaud, E. Bandman and R. Ferrand. 1992. Developmental modulation of myosin expression by thyroid hormone in avian skeletal muscle. *Development. 115*: 1121-1131.

George-Weinstein, M., J.V. Gerhart. G.J. Foti and J.W. Lash. 1994. Maturation of myogenic and chondrogenic cells in the presomitic mesoderm of the chick embryo. *Exp. Cell Res. 211*: 263-274.

Goulding, M., A. Lumsden and A.J. Paquette. 1994. Regulation of Pax-3 expression in the dermamyotome and its role in muscle development. *Development. 120*: 957-971.

Halpern, M.E., R.K. Ho, C. Walker and C.B. Kimmel. 1993. Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell. 75*: 99-111.

Halpern, M.E., C. Thisse. R.K. Ho. B. Thisse. B. Riggleman. B. Trevarrow, E.S. Weinberg, J.H. Postlethwait and C.B. Kimmel. 1995. Cell-autonomous shift from axial to paraxial mesodermal development in zebrafish floating head mutants. *Development. 121*: 4257-4264.

Hammerschmidt, M., M.J. Bitgood and A.P. McMahon. 1996. Protein kinase A is a common negative regulator of hedgehog signalling in the vertebrate embryo. *Genes Dev.* ***10***: 647-658.

Hatta, K., R. Bremiller, M. Westerfield and C.B. Kimmel. 1991. Diversity of expression of *engrailed*-like antigens in zebrafish. *Development.* ***112***: 821-832.

Hatta, K., T.F. Schilling, R.A. BreMiller and C.B. Kimmel. 1990. Specification of jaw muscle identity in zebrafish: correlation with *engrailed*-homeoprotein expression. *Science.* ***250***: 802-805.

Holland, L.Z., D.A. Pace, M.L. Blink, M. Kene and N.D. Holland. 1995. Sequence and expression of amphioxus alkali myosin light chain (AmphiMLC-alk) throughout development: implications for vertebrate myogenesis. *Dev.Biol.* ***171***: 665-676.

Huang, Z. and S. Kunes. 1996. Hedgehog, transmitted along retinal axons, triggers neurogensis in the developing visual centers of the Drosophila brain. *Cell.* ***86***: 411-422.

Hughes, S.M., M. Cho, I. Karsch-Mizrachi, M. Travis, L. Silberstein, L.A. Leinwand and H.M. Blau. 1993. Three slow myosin heavy chains sequentially expressed in developing mammalian skeletal muscle. *Dev. Biol.* ***158***: 183-199.

Hughes, S.M. and H.M. Blau. 1992. Muscle fiber pattern is independent of cell lineage in postnatal rodent development. *Cell.* ***68***: 659-671.

Johnson, R.L., E. Laufer, R.D. Riddle and C. Tabin. 1994. Ectopic expression of Sonic hedgehog alters dorsal-ventral patterning of somites. *Cell.* ***79***: 1165-1173.

Kelly, A.M. and N.A. Rubinstein. 1980. Why are fetal muscles slow? *Nature.* ***288***: 266-269.

Kenny-Mobbs, T. and P. Thorogood. 1987. Autonomy of differentiation in avian brachial somites and the influence of adjacent tissues. *Development.* ***100***: 449-462.

Kimmel, C.B. and R.M. Warga. 1987. Cell lineages generating axial muscle in the zebrafish embryo. *Nature.* ***327***: 234-237.

Krauss, S., J.-P. Concordet and P.W. Ingham. 1993. A functionally conserved homologue of the *Drosophila* segment polarity gene *hh* is expressed in tissues with polarizing activity in zebrafish embryos. *Cell. 75*: 1431-1444.

Lassar, A.B. and A.E. Munsterberg. 1996. The role of positive and negative signals in somite patterning. *Curr.Opinions Neurobiol. 6*: 57-63.

~~Lin-Jones, J. and S.D. Hauschka. 1996. Myogenic determination factor expression in the~~ developing avian limb bud: an RT-PCR analysis. *Dev.Biol. 174*: 407-422.

Marigo, V. and C.J. Tabin. 1996. Regulation of *patched* by sonic hedgehog in the developing neural tube. *Proc. Natl. Acad. Sci. USA. 93*: 9346-9351.

Maroto, M., R. Reshef, A.E. Münsterberg, S. Koester, M. Goulding and A.B. Lassar. 1997. Ectopic *Pax-3* activates *MyoD* and *Myf-5* expression in embryonic mesoderm and neural tissue. *Cell* 89, 139-148.

Marti, E., R. Takada, D.A. Bumcrot, H. Sasaki and A.P. McMahon. 1995. Distribution of Sonic hedgehog peptides in the developing chick and mouse embryo. *Development. 121*: 2537-2547.

Miller, J.B. and F.E. Stockdale. 1986. Developmental origins of skeletal muscle fibres: clonal analysis of myogenic cell lineages based on expression of fast and slow myosin heavy chains. *Proc.Natl. Acad.Sci. USA. 83*: 3860-3864.

Morin-Kensicki, E.M. and J.S. Eisen. 1997. Sclerotome development and peripheral nervous system segmentation in embryonic zebrafish. *Development. 124*: 159-167.

Munsterberg, A.E., J. Kitajewski, D.A. Bumcrot, A.P. McMahon and A.B. Lassar. 1995. Combinatorial signalling by sonic hedgehog and Wnt family members induces myogenic bHLH gene expression in the somite. *Genes & Dev. 9*: 2911-2922.

Odenthal, J., P. Haffter, E. Vogelsang, M. Brand, F.J.M. Van Eeden, M. Furutani-Seiki, M. Granato, M. Hammerschmidt, C.P. Heisenberg, Y.J. Jiang, D.A. Kane, R.N. Kelsh, M.C. Mullins, R.M. Warga, M.L. Allende, E.S. Weinberg and C. NussleinVolhard. 1996.

Mutations affecting the formation of the notochord in the zebrafish. Danio rerio. *Development.* *123*: 103-115.

Ott. M.-O.. E. Bober. G. Lyons. H. Arnold and M. Buckingham. 1991. Early expression of the myogenic regulatory gene. *myf-5*. in precursor cells of skeletal muscle in the mouse embryo. *Development.* *111*: 1097-1107.

Page. S.. J.B. Miller. J.X. DiMario. E.J. Hager. A. Moser and F.E. Stockdale. 1992. Developmentally regulated expression of three slow isoforms of myosin heavy chain: Diversity among the first fibers to form in avian muscle. *Dev. Biol.* 154. 118-128.

Pownall. M.E.. K.E. Strunk and C.P. Emerson. Jr. 1996. Notochord signals control the transcriptional cascade of myogenic bHLH genes in somites of quail embryos. *Development* 122. 1475-1488.

Pourquie. O.. M. Coltey. M.-A. Teillet. C. Ordahl and N.M. Le Douarin. 1993. Control of dorsoventral patterning of somitic derivatives by notochord and floor plate. *Proc.Natl.Acad.Sci. USA.* *90*: 5242-5246.

Pourquie. O.. C.-M. Fan. M. Coltey. E. Hirsinger. Y. Watanabe. C. Breant. P. Francis-West. P. Brickell. M. Tessier-Lavigne and N.M. Le Douarin. 1996. Lateral and axial signals involved in avian somite patterning: a role for BMP4. *Cell.* *84*: 461-471.

Quinn. L.S.. H. Holtzer and M. Nameroff. 1985. Generation of chick skeletal muscle cells in groups of 16 from stem cells. *Nature.* *313*: 692-694.

Roelink. H.. A. Augsburger. J. Heemskerk. V. Korzh. S. Norlin. A. Ruiz i Altaba. Y.. Tanabe. M. Placzek. T. Edlund. T.M. Jessell and J. Dodd. 1994. Floor plate and motor neuron induction by *vhh-1*. a vertebrate homologue of *hedgehog* expressed by the notochord. *Cell.* *76*: 761-775.

Rong. P.M.. A.-M. Teillet. C. Ziller and N.M. Le Douarin. 1992. The neural tube/notochord complex is necessary for vertebral but not limb and body wall striated muscle differentiation. *Development.* *115*: 657-672.

Rudnicki, M.A., P.N.J. Schnegelsberg, R.H. Stead, T. Braun, H.-H. Arnold and R. Jaenisch. 1993. MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell. 75*: 1351-1359.

Schafer, D.A., J.B. Miller and F.E. Stockdale. 1987. Cell diversification within the myogenic lineage: in vitro generation of two types of myoblasts from a single myogenic progenitor cell. *Cell 48*: 659-670.

Schier, A.F., S.C.F. Neuhauss, K.A. Helde, W.S. Talbot and W. Driever. 1997. The *one-eyed pinhead* gene functions in mesoderm and endoderm formation in zebrafish and interacts with *no tail*. *Development. 124*: 327-342.

Schiaffino, S., L. Gorza, S. Sartore, L. Saggin, S. Ausoni, M. Vianello, K. Gundersen and T. Lomo. 1989. Three myosin heavy chain isoforms in type 2 skeletal muscle fibres. *J.Muscle Res. Cell Motil. 10*: 197-205.

Schulte-Merker, S., F.J.M. van Eeden, M.E. Halpern, C.B. Kimmel and C. Nusslein-Volhard. 1994. *no tail (ntl)* is the zebrafish homologue of the mouse *T (Brachyury)* gene. *Development. 120*: 1009-1015.

Solnica-Krezel, L., D.L. Stemple, E. Mountcastle-Shah, Z. Rangini, S.C.F. Neuhauss, J. Malicki, A.F. Schier, D.Y.R. Stainier, F. Zwartkruis, S. Abdelilah and W. Driever. 1996. Mutations affecting cell fates and cellular rearrangements during gastrulation in zebrafish. *Development. 123*: 67-80.

Stern, H.M., A.M.C. Brown and S.D. Hauschka. 1995. Myogenesis in paraxial mesoderm: Preferential induction by dorsal neural tube and by cells expressing Wnt-1. *Development. 121*: 3675-3686.

Stern, H.M. and S.D. Hauschka. 1995. Neural tube and notochord promote *in vitro* myogenesis in single somite explants. *Dev. Biol. 167*: 87-103.

Stone, D.M., M. Hynes, M. Armanini, T.A. Swanson, Q. Gu, R.L. Johnson, M.P. Scott, D. Pennica, A. Goddard, H. Phillips, M. Noll, J.E. Hooper, F. De Sauvage and A. Rosenthal.

1996. The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature. 384*: 129-134.

Tajbakhsh, S., D. Rocancourt, G. Cossu and M. Buckingham. 1997. Redefining the genetic hierarchies controlling skeletal myogenesis: Pax-3 and Myf-5 act upstream of MyoD. *Cell* 89, 127-138.

Talbot, W.B., S. Trevärrow, M.E. Halpern, A.E. Melby, G. Farr, T.J. Postlethwait, C.B. Kimmel and D. Kimmelman. 1995. A homeobox gene essential for zebrafish notochord development. *Nature. 378*: 150-157.

Teillet, M.-A. and N.M. Le Douarin. 1983. Consequences of neural tube and notochord excision on the development of the peripheral nervous system in chick embryos. *Dev. Biol. 98*: 192-211.

Ticho, B.S., D.Y.R. Stainier and M.C. Fishman. 1996. Three zebrafish MEF2 genes delineate somitic and cardiac muscle development in wild-type and mutant embryos. *Mech. Dev. 59*: 205-218.

van Eeden, F.J.M., M. Granato, U. Schach, M. Brand, M. Furutani-Seiki, P. Haffter, M. Hammerschmidt, C.P. Heisenberg, Y.-J. Jiang, D.A. Kane, R.N. Kelsh, M.C. Mullins, J. Odenthal, R.M. Warga, M.L. Allende, E.S. Weinberg and C. Nusslein-Volhard. 1996. Mutations affecting somite formation and patterning in the zebrafish, Danio rerio. *Development* 123, 153-164.

van Raamsdonk, W., C.W. Pool and G. te Kronnie. 1978. Differentiation of muscle fibre types in the teleost *Brachydanio rerio*. *Anat. Embryol. 153*: 137-155.

Vivarelli, E., W.E. Brown, R.G. Whalen and G. Cossu. 1988. The expression of slow myosin during mammalian somitogenesis and limb bud differentiation. *J. Cell Biol. 107*: 2191-2197.

Vortkamp, A., K. Lee, B. Lanske, G.V. Segre, H.M. Kronenberg and C.J. Tabin. 1996. Regulation of rate of cartilage differentiation by Indian Hedgehog and PTH-related protein. *Science. 273*: 613-622.

Webster, C., L. Silberstein, A.P. Hays and H.M. Blau. 1988. Fast muscle fibers are preferentially affected in Duchenne muscular dystrophy. *Cell*. *52*: 503-513.

Weinberg, E.S., M.L. Allende, C.S. Kelly, A. Abdelhamid, T. Murakami, P. Andermann, O.G. Doerre, D.J. Grunwald and B. Riggleman. 1996. Developmental regulation of zebrafish MyoD in wild-type, no tail and spadetail embryos. *Development*. *122*: 271-280.

All of the above-cited references and publications are hereby incorporated by reference.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, numerous equivalents to the specific polypeptides, nucleic acids, methods, assays and reagents described herein. Such equivalents are considered to be within the scope of this invention.

Claims

1. A method for regulating formation and/or maintenance of muscle tissue comprising contacting the muscle cells with a hedgehog polypeptide or a *ptc* therapeutic.

2. The method of claim 1, wherein the hedgehog polypeptide is modified with one or more lipophilic moieties.

3. The method of claim 2, wherein the hedgehog polypeptide is modified with one or more serol moieties.

4. The method of claim 2, wherein the sterol moiety is cholesterol.

5. The method of claim 2, wherein the hedgehog polypeptide is modified with one or more fatty acid moieties.

6. The method of claim 5, wherein each fatty acid moiety is independently selected from the group consisting of myristoyl, palmitoyl, stearoyl, and arachidoyl.

7. The method of claim 2, wherein the hedgehog polypeptide is modified with one or more aromatic hydrocarbons.

8. The method of claim 1, wherein the *ptc* therapeutic binds to *patched* and mimics *hedgehog*-mediated *patched* signal transduction.

9. The method of claim 8, wherein the *ptc* therapeutic is a small organic molecule.

10. The method of claim 8, wherein the binding of the ptc therapeutic to *patched* results in upregulation of patched and/or gli expression.

11. The method of claim 1, wherein the *ptc* therapeutic is a small organic molecule which interacts with neuronal cells to mimic *hedgehog*-mediated *patched* signal transduction.

12. The method of claim 1, wherein the *ptc* therapeutic mimics *hedgehog*-mediated *patched* signal transduction by altering the localization, protein-protein binding and/or enzymatic activity of an intracellular protein involved in a *patched* signal pathway.

13. The method of claim 1, wherein the *ptc* therapeutic alters the level of expression of a *hedgehog* protein, a patched protein or a protein involved in the intracellular signal transduction pathway of *patched*.

14. The method of claim 13, wherein the *ptc* therapeutic is an antisense construct which inhibits the expression of a protein which is involved in the signal

transduction pathway of *patched* and the expression of which antagonizes *hedgehog*-mediated signals.

15. The method of claim 14, wherein the antisense construct is an oligonucleotide of about 20-30 nucleotides in length and having a GC content of at least 50 percent.

16. The method of claim 15, wherein the antisense oligonucleotide is selected from the group consisting of: 5'-GTCCTGGCGCCGCCGCCGCCGTCGCC;

5'-TTCCGATGACCGGCCTTTCGCGGTGA; and

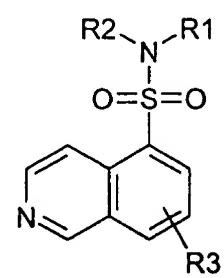
5'-GTGCACGGAAAGGTGCAGGCCACACT

17. The method of claims 13, wherein the *ptc* therapeutic is a small organic molecule which binds to *patched* and regulates *patched*-dependent gene expression.

18. The method of claim 12, wherein the *ptc* therapeutic is an inhibitor of protein kinase A.

19. The method of claim 18, wherein the PKA inhibitor is a 5-isoquinolinesulfonamide

20. The method of claim 19, wherein the PKA inhibitor is represented in the general formula:



wherein,

R_1 and R_2 each can independently represent hydrogen, and as valence and stability permit a lower alkyl, a lower alkenyl, a lower alkynyl, a carbonyl (such as a carboxyl, an ester, a formate, or a ketone), a thiocarbonyl (such as a thioester, a thioacetate, or a thioformate), an amino, an acylamino, an amido, a cyano, a nitro, an azido, a sulfate, a sulfonate, a sulfonamido, $-(CH_2)_m-R_8$, $-(CH_2)_m-OH$, $-(CH_2)_m-O-$lower alkyl, $-(CH_2)_m-O-$lower alkenyl, $-(CH_2)_n-O-(CH_2)_m-R_8$, $-(CH_2)_m-SH$, $-(CH_2)_m-S-$lower alkyl, $-(CH_2)_m-S-$lower alkenyl, $-(CH_2)_n-S-(CH_2)_m-R_8$, or

R_1 and R_2 taken together with N form a heterocycle (substituted or unsubstituted);

R_3 is absent or represents one or more substitutions to the isoquinoline ring such as a lower alkyl, a lower alkenyl, a lower alkynyl, a carbonyl (such as a carboxyl, an ester, a formate, or a ketone), a thiocarbonyl (such as a thioester, a thioacetate, or a thioformate), an amino, an acylamino, an amido, a cyano, a nitro, an azido, a sulfate, a sulfonate, a sulfonamido, $-(CH_2)_m-R_8$, $-(CH_2)_m-OH$, $-(CH_2)_m$-O-lower alkyl, $-(CH_2)_m$-O-lower alkenyl, $-(CH_2)_n-O-(CH_2)_m-R_8$, $-(CH_2)_m-SH$, $-(CH_2)_m$-S-lower alkyl, $-(CH_2)_m$-S-lower alkenyl, $-(CH_2)_n-S-(CH_2)_m-R_8$;

R_8 represents a substituted or unsubstituted aryl, aralkyl, cycloalkyl, cycloalkenyl, or heterocycle; and

n and m are independently for each occurrence zero or an integer in the range of 1 to 6.

21. The method of claim 18, wherein the PKA inhibitor is cyclic AMP analog.

22. The method of claim 18, wherein the PKA inhibitor is selected from the group consisting of N-[2-((p-bromocinnamyl)amino)ethyl]-5-isoquinolinesulfonamide, 1-(5-isoquinoline-sulfonyl)-2-methylpiperazine, KT5720, 8-bromo-cAMP, dibutyryl-cAMP and PKA Heat Stable Inhibitor isoform α.

23. The method of claim 1, wherein patient is being treated prophylactically.

24. A therapeutic preparation of a small molecule antagonist of *patched*, which *patched* antagonist is provided in a pharmaceutically acceptable carrier and in an amount sufficient to regulate growth and/or maintenance of muscle cells.

25. A method for regulating the growth state of a muscle stem/progenitor cell comprising contacting the cell with a hedgehog polypeptide or a *ptc* therapeutic.

26. A method for treatment or prevention of disorders of, or surgical or cosmetic repair of, such muscle tissues, comprising administering to the patient a hedgehog polypeptide or a *ptc* therapeutic.

27. The method of claim 26, wherein the disorder is muscle atrophy.

28. The method of claim 27, wherein the disorder is skeletal muscle atrophy or cardiac muscle atrophy.

29. The method of claim 26, wherein the disorder is cachexia.

30. The method of claim 26, wherein the disorder is a muscular myopathy.

31. The method of claim 26, wherein the hedgehog polypeptide or *ptc* therapeutic administered inhibits growth of myoblastic-derived tissue, and disorder is includes hyperplastic or neoplastic growth of muscle tissue.

32. The method of claim 40, wherein the disorder is myoblastic sarcoma.

SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO:1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1277 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1275

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG GTC GAA ATG CTG CTG TTG ACA AGA ATT CTC TTG GTG GGC TTC ATC       48
Met Val Glu Met Leu Leu Leu Thr Arg Ile Leu Leu Val Gly Phe Ile
  1               5                  10                  15

TGC GCT CTT TTA GTC TCC TCT GGG CTG ACT TGT GGA CCA GGC AGG GGC       96
Cys Ala Leu Leu Val Ser Ser Gly Leu Thr Cys Gly Pro Gly Arg Gly
             20                  25                  30

ATT GGA AAA AGG AGG CAC CCC AAA AAG CTG ACC CCG TTA GCC TAT AAG      144
Ile Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys
         35                  40                  45

CAG TTT ATT CCC AAT GTG GCA GAG AAG ACC CTA GGG GCC AGT GGA AGA      192
Gln Phe Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg
     50                  55                  60

TAT GAA GGG AAG ATC ACA AGA AAC TCC GAG AGA TTT AAA GAA CTA ACC      240
Tyr Glu Gly Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr
 65                  70                  75                  80

CCA AAT TAC AAC CCT GAC ATT ATT TTT AAG GAT GAA GAG AAC ACG GGA      288
Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly
                 85                  90                  95

GCT GAC AGA CTG ATG ACT CAG CGC TGC AAG GAC AAG CTG AAT GCC CTG      336
Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu
            100                 105                 110

GCG ATC TCG GTG ATG AAC CAG TGG CCC GGG GTG AAG CTG CGG GTG ACC      384
Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr
        115                 120                 125

GAG GGC TGG GAC GAG GAT GGC CAT CAC TCC GAG GAA TCG CTG CAC TAC      432
Glu Gly Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr
    130                 135                 140

GAG GGT CGC GCC GTG GAC ATC ACC ACG TCG GAT CGG GAC CGC AGC AAG      480
Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys
145                 150                 155                 160

TAC GGA ATG CTG GCC CGC CTC GCC GTC GAG GCC GGC TTC GAC TGG GTC      528
```

```
Tyr Gly Met Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val
                165                 170                 175

TAC TAC GAG TCC AAG GCG CAC ATC CAC TGC TCC GTC AAA GCA GAA AAC      576
Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn
            180                 185                 190

TCA GTG GCA GCG AAA TCA GGA GGC TGC TTC CCT GGC TCA GCC ACA GTG      624
Ser Val Ala Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Thr Val
        195                 200                 205

CAC CTG GAG CAT GGA GGC ACC AAG CTG GTG AAG GAC CTG AGC CCT GGG      672
His Leu Glu His Gly Gly Thr Lys Leu Val Lys Asp Leu Ser Pro Gly
    210                 215                 220

GAC CGC GTG CTG GCT GCT GAC GCG GAC GGC CGG CTG CTC TAC AGT GAC      720
Asp Arg Val Leu Ala Ala Asp Ala Asp Gly Arg Leu Leu Tyr Ser Asp
225                 230                 235                 240

TTC CTC ACC TTC CTC GAC CGG ATG GAC AGC TCC CGA AAG CTC TTC TAC      768
Phe Leu Thr Phe Leu Asp Arg Met Asp Ser Ser Arg Lys Leu Phe Tyr
                245                 250                 255

GTC ATC GAG ACG CGG CAG CCC CGG GCC CGG CTG CTA CTG ACG GCG GCC      816
Val Ile Glu Thr Arg Gln Pro Arg Ala Arg Leu Leu Leu Thr Ala Ala
            260                 265                 270

CAC CTG CTC TTT GTG GCC CCC CAG CAC AAC CAG TCG GAG GCC ACA GGG      864
His Leu Leu Phe Val Ala Pro Gln His Asn Gln Ser Glu Ala Thr Gly
        275                 280                 285

TCC ACC AGT GGC CAG GCG CTC TTC GCC AGC AAC GTG AAG CCT GGC CAA      912
Ser Thr Ser Gly Gln Ala Leu Phe Ala Ser Asn Val Lys Pro Gly Gln
    290                 295                 300

CGT GTC TAT GTG CTG GGC GAG GGC GGG CAG CAG CTG CTG CCG GCG TCT      960
Arg Val Tyr Val Leu Gly Glu Gly Gly Gln Gln Leu Leu Pro Ala Ser
305                 310                 315                 320

GTC CAC AGC GTC TCA TTG CGG GAG GAG GCG TCC GGA GCC TAC GCC CCA     1008
Val His Ser Val Ser Leu Arg Glu Glu Ala Ser Gly Ala Tyr Ala Pro
                325                 330                 335

CTC ACC GCC CAG GGC ACC ATC CTC ATC AAC CGG GTG TTG GCC TCC TGC     1056
Leu Thr Ala Gln Gly Thr Ile Leu Ile Asn Arg Val Leu Ala Ser Cys
            340                 345                 350

TAC GCC GTC ATC GAG GAG CAC AGT TGG GCC CAT TGG GCC TTC GCA CCA     1104
Tyr Ala Val Ile Glu Glu His Ser Trp Ala His Trp Ala Phe Ala Pro
        355                 360                 365

TTC CGC TTG GCT CAG GGG CTG CTG GCC GCC CTC TGC CCA GAT GGG GCC     1152
Phe Arg Leu Ala Gln Gly Leu Leu Ala Ala Leu Cys Pro Asp Gly Ala
    370                 375                 380

ATC CCT ACT GCC GCC ACC ACC ACC ACT GGC ATC CAT TGG TAC TCA CGG     1200
Ile Pro Thr Ala Ala Thr Thr Thr Thr Gly Ile His Trp Tyr Ser Arg
385                 390                 395                 400

CTC CTC TAC CGC ATC GGC AGC TGG GTG CTG GAT GGT GAC GCG CTG CAT     1248
Leu Leu Tyr Arg Ile Gly Ser Trp Val Leu Asp Gly Asp Ala Leu His
                405                 410                 415
```

```
CCG CTG GGC ATG GTG GCA CCG GCC AGC TG                          1277
Pro Leu Gly Met Val Ala Pro Ala Ser
            420                 425
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1190 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1191

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
ATG GCT CTG CCG GCC AGT CTG TTG CCC CTG TGC TGC TTG GCA CTC TTG       48
Met Ala Leu Pro Ala Ser Leu Leu Pro Leu Cys Cys Leu Ala Leu Leu
  1               5                  10                  15

GCA CTA TCT GCC CAG AGC TGC GGG CCG GGC CGA GGA CCG GTT GGC CGG       96
Ala Leu Ser Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val Gly Arg
             20                  25                  30

CGG CGT TAT GTG CGC AAG CAA CTT GTG CCT CTG CTA TAC AAG CAG TTT      144
Arg Arg Tyr Val Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe
         35                  40                  45

GTG CCC AGT ATG CCC GAG CGG ACC CTG GGC GCG AGT GGG CCA GCG GAG      192
Val Pro Ser Met Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu
     50                  55                  60

GGG AGG GTA ACA AGG GGG TCG GAG CGC TTC CGG GAC CTC GTA CCC AAC      240
Gly Arg Val Thr Arg Gly Ser Glu Arg Phe Arg Asp Leu Val Pro Asn
 65                  70                  75                  80

TAC AAC CCC GAC ATA ATC TTC AAG GAT GAG GAG AAC AGC GGC GCA GAC      288
Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp
                 85                  90                  95

CGC CTG ATG ACA GAG CGT TGC AAA GAG CGG GTG AAC GCT CTA GCC ATC      336
Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile
            100                 105                 110

GCG GTG ATG AAC ATG TGG CCC GGA GTA CGC CTA CGT GTG ACT GAA GGC      384
Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly
        115                 120                 125

TGG GAC GAG GAC GGC CAC CAC GCA CAG GAT TCA CTC CAC TAC GAA GGC      432
Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly
    130                 135                 140

CGT GCC TTG GAC ATC ACC ACG TCT GAC CGT GAC CGT AAT AAG TAT GGT      480
```

```
Arg Ala Leu Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly
145                 150                 155                 160

TTG TTG GCG CGC CTA GCT GTG GAA GCC GGA TTC GAC TGG GTC TAC TAC      528
Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
                165                 170                 175

GAG TCC CGC AAC CAC ATC CAC GTA TCG GTC AAA GCT GAT AAC TCA CTG      576
Glu Ser Arg Asn His Ile His Val Ser Val Lys Ala Asp Asn Ser Leu
            180                 185                 190

GCG GTC CGA GCC GGA GGC TGC TTT CCG GGA AAT GCC ACG GTG CGC TTG      624
Ala Val Arg Ala Gly Gly Cys Phe Pro Gly Asn Ala Thr Val Arg Leu
        195                 200                 205

CGG AGC GGC GAA CGG AAG GGG CTG AGG GAA CTA CAT CGT GGT GAC TGG      672
Arg Ser Gly Glu Arg Lys Gly Leu Arg Glu Leu His Arg Gly Asp Trp
    210                 215                 220

GTA CTG GCC GCT GAT GCA GCG GGC CGA GTG GTA CCC ACG CCA GTG CTG      720
Val Leu Ala Ala Asp Ala Ala Gly Arg Val Val Pro Thr Pro Val Leu
225                 230                 235                 240

CTC TTC CTG GAC CGG GAT CTG CAG CGC CGC GCC TCG TTC GTG GCT GTG      768
Leu Phe Leu Asp Arg Asp Leu Gln Arg Arg Ala Ser Phe Val Ala Val
                245                 250                 255

GAG ACC GAG CGG CCT CCG CGC AAA CTG TTG CTC ACA CCC TGG CAT CTG      816
Glu Thr Glu Arg Pro Pro Arg Lys Leu Leu Leu Thr Pro Trp His Leu
            260                 265                 270

GTG TTC GCT GCT CGC GGG CCA GCG CCT GCT CCA GGT GAC TTT GCA CCG      864
Val Phe Ala Ala Arg Gly Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro
        275                 280                 285

GTG TTC GCG CGC CGC TTA CGT GCT GGC GAC TCG GTG CTG GCT CCC GGC      912
Val Phe Ala Arg Arg Leu Arg Ala Gly Asp Ser Val Leu Ala Pro Gly
    290                 295                 300

GGG GAC GCG CTC CAG CCG GCG CGC GTA GCC CGC GTG GCG CGC GAG GAA      960
Gly Asp Ala Leu Gln Pro Ala Arg Val Ala Arg Val Ala Arg Glu Glu
305                 310                 315                 320

GCC GTG GGC GTG TTC GCA CCG CTC ACT GCG CAC GGG ACG CTG CTG GTC     1008
Ala Val Gly Val Phe Ala Pro Leu Thr Ala His Gly Thr Leu Leu Val
                325                 330                 335

AAC GAC GTC CTC GCC TCC TGC TAC GCG GTT CTA GAG AGT CAC CAG TGG     1056
Asn Asp Val Leu Ala Ser Cys Tyr Ala Val Leu Glu Ser His Gln Trp
            340                 345                 350

GCC CAC CGC GCC TTC GCC CCT TTG CGG CTG CTG CAC GCG CTC GGG GCT     1104
Ala His Arg Ala Phe Ala Pro Leu Arg Leu Leu His Ala Leu Gly Ala
        355                 360                 365

CTG CTC CCT GGG GGT GCA GTC CAG CCG ACT GGC ATG CAT TGG TAC TCT     1152
Leu Leu Pro Gly Gly Ala Val Gln Pro Thr Gly Met His Trp Tyr Ser
    370                 375                 380

CGC CTC CTT TAC CGC TTG GCC GAG GAG TTA ATG GGC TGA                 1191
Arg Leu Leu Tyr Arg Leu Ala Glu Glu Leu Met Gly
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1281 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1233

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG TCT CCC GCC TGG CTC CGG CCC CGA CTG CGG TTC TGT CTG TTC CTG       48
Met Ser Pro Ala Trp Leu Arg Pro Arg Leu Arg Phe Cys Leu Phe Leu
 1               5                  10                  15

CTG CTG CTG CTT CTG GTG CCG GCG GCG CGG GGC TGC GGG CCG GGC CGG       96
Leu Leu Leu Leu Leu Val Pro Ala Ala Arg Gly Cys Gly Pro Gly Arg
             20                  25                  30

GTG GTG GGC AGC CGC CGG AGG CCG CCT CGC AAG CTC GTG CCT CTT GCC      144
Val Val Gly Ser Arg Arg Arg Pro Pro Arg Lys Leu Val Pro Leu Ala
         35                  40                  45

TAC AAG CAG TTC AGC CCC AAC GTG CCG GAG AAG ACC CTG GGC GCC AGC      192
Tyr Lys Gln Phe Ser Pro Asn Val Pro Glu Lys Thr Leu Gly Ala Ser
     50                  55                  60

GGG CGC TAC GAA GGC AAG ATC GCG CGC AGC TCT GAG CGC TTC AAA GAG      240
Gly Arg Tyr Glu Gly Lys Ile Ala Arg Ser Ser Glu Arg Phe Lys Glu
 65                  70                  75                  80

CTC ACC CCC AAC TAC AAT CCC GAC ATC ATC TTC AAG GAC GAG GAG AAC      288
Leu Thr Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn
                 85                  90                  95

ACG GGT GCC GAC CGC CTC ATG ACC CAG CGC TGC AAG GAC CGT CTG AAC      336
Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Arg Leu Asn
            100                 105                 110

TCA CTG GCC ATC TCT GTC ATG AAC CAG TGG CCT GGT GTG AAA CTG CGG      384
Ser Leu Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg
        115                 120                 125

GTG ACC GAA GGC CGG GAT GAA GAT GGC CAT CAC TCA GAG GAG TCT TTA      432
Val Thr Glu Gly Arg Asp Glu Asp Gly His His Ser Glu Glu Ser Leu
    130                 135                 140

CAC TAT GAG GGC CGC GCG GTG GAT ATC ACC ACC TCA GAC CGT GAC CGA      480
His Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg
145                 150                 155                 160

AAT AAG TAT GGA CTG CTG GCG CGC TTA GCA GTG GAG GCC GGC TTC GAC      528
Asn Lys Tyr Gly Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp
```

```
                165                 170                 175

TGG GTG TAT TAC GAG TCC AAG GCC CAC GTG CAT TGC TCT GTC AAG TCT      576
Trp Val Tyr Tyr Glu Ser Lys Ala His Val His Cys Ser Val Lys Ser
            180                 185                 190

GAG CAT TCG GCC GCT GCC AAG ACA GGT GGC TGC TTT CCT GCC GGA GCC      624
Glu His Ser Ala Ala Ala Lys Thr Gly Gly Cys Phe Pro Ala Gly Ala
        195                 200                 205

CAG GTG CGC CTA GAG AAC GGG GAG CGT GTG GCC CTG TCA GCT GTA AAG      672
Gln Val Arg Leu Glu Asn Gly Glu Arg Val Ala Leu Ser Ala Val Lys
    210                 215                 220

CCA GGA GAC CGG GTG CTG GCC ATG GGG GAG GAT GGG ACC CCC ACC TTC      720
Pro Gly Asp Arg Val Leu Ala Met Gly Glu Asp Gly Thr Pro Thr Phe
225                 230                 235                 240

AGT GAT GTG CTT ATT TTC CTG GAC CGC GAG CCA AAC CGG CTG AGA GCT      768
Ser Asp Val Leu Ile Phe Leu Asp Arg Glu Pro Asn Arg Leu Arg Ala
                245                 250                 255

TTC CAG GTC ATC GAG ACT CAG GAT CCT CCG CGT CGG CTG GCG CTC ACG      816
Phe Gln Val Ile Glu Thr Gln Asp Pro Pro Arg Arg Leu Ala Leu Thr
            260                 265                 270

CCT GCC CAC CTG CTC TTC ATT GCG GAC AAT CAT ACA GAA CCA GCA GCC      864
Pro Ala His Leu Leu Phe Ile Ala Asp Asn His Thr Glu Pro Ala Ala
        275                 280                 285

CAC TTC CGG GCC ACA TTT GCC AGC CAT GTG CAA CCA GGC CAA TAT GTG      912
His Phe Arg Ala Thr Phe Ala Ser His Val Gln Pro Gly Gln Tyr Val
    290                 295                 300

CTG GTA TCA GGG GTA CCA GGC CTC CAG CCT GCT CGG GTG GCA GCT GTC      960
Leu Val Ser Gly Val Pro Gly Leu Gln Pro Ala Arg Val Ala Ala Val
305                 310                 315                 320

TCC ACC CAC GTG GCC CTT GGG TCC TAT GCT CCT CTC ACA AGG CAT GGG     1008
Ser Thr His Val Ala Leu Gly Ser Tyr Ala Pro Leu Thr Arg His Gly
                325                 330                 335

ACA CTT GTG GTG GAG GAT GTG GTG GCC TCC TGC TTT GCA GCT GTG GCT     1056
Thr Leu Val Val Glu Asp Val Val Ala Ser Cys Phe Ala Ala Val Ala
            340                 345                 350

GAC CAC CAT CTG GCT CAG TTG GCC TTC TGG CCC CTG CGA CTG TTT CCC     1104
Asp His His Leu Ala Gln Leu Ala Phe Trp Pro Leu Arg Leu Phe Pro
        355                 360                 365

AGT TTG GCA TGG GGC AGC TGG ACC CCA AGT GAG GGT GTT CAC TCC TAC     1152
Ser Leu Ala Trp Gly Ser Trp Thr Pro Ser Glu Gly Val His Ser Tyr
    370                 375                 380

CCT CAG ATG CTC TAC CGC CTG GGG CGT CTC TTG CTA GAA GAG AGC ACC     1200
Pro Gln Met Leu Tyr Arg Leu Gly Arg Leu Leu Leu Glu Glu Ser Thr
385                 390                 395                 400

TTC CAT CCA CTG GGC ATG TCT GGG GCA GGA AGC TGAAGGGACT CTAACCACTG   1253
Phe His Pro Leu Gly Met Ser Gly Ala Gly Ser
                405                 410
```

```
CCCTCCTGGA ACTGCTGTGC GTGGATCC                                        1281
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1313 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1314

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
ATG CTG CTG CTG CTG GCC AGA TGT TTT CTG GTG ATC CTT GCT TCC TCG       48
Met Leu Leu Leu Leu Ala Arg Cys Phe Leu Val Ile Leu Ala Ser Ser
  1               5                  10                  15

CTG CTG GTG TGC CCC GGG CTG GCC TGT GGG CCC GGC AGG GGG TTT GGA       96
Leu Leu Val Cys Pro Gly Leu Ala Cys Gly Pro Gly Arg Gly Phe Gly
             20                  25                  30

AAG AGG CGG CAC CCC AAA AAG CTG ACC CCT TTA GCC TAC AAG CAG TTT      144
Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe
         35                  40                  45

ATT CCC AAC GTA GCC GAG AAG ACC CTA GGG GCC AGC GGC AGA TAT GAA      192
Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg Tyr Glu
     50                  55                  60

GGG AAG ATC ACA AGA AAC TCC GAA CGA TTT AAG GAA CTC ACC CCC AAT      240
Gly Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn
 65                  70                  75                  80

TAC AAC CCC GAC ATC ATA TTT AAG GAT GAG GAA AAC ACG GGA GCA GAC      288
Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp
                 85                  90                  95

CGG CTG ATG ACT CAG AGG TGC AAA GAC AAG TTA AAT GCC TTG GCC ATC      336
Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala Ile
            100                 105                 110

TCT GTG ATG AAC CAG TGG CCT GGA GTG AGG CTG CGA GTG ACC GAG GGC      384
Ser Val Met Asn Gln Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly
        115                 120                 125

TGG GAT GAG GAC GGC CAT CAT TCA GAG GAG TCT CTA CAC TAT GAG GGT      432
Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr Glu Gly
    130                 135                 140

CGA GCA GTG GAC ATC ACC ACG TCC GAC CGG GAC CGC AGC AAG TAC GGC      480
Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly
145                 150                 155                 160

ATG CTG GCT CGC CTG GCT GTG GAA GCA GGT TTC GAC TGG GTC TAC TAT      528
```

```
Met Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
                165                 170                 175

GAA TCC AAA GCT CAC ATC CAC TGT TCT GTG AAA GCA GAG AAC TCC GTG      576
Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val
            180                 185                 190

GCG GCC AAA TCC GGC GGC TGT TTC CCG GGA TCC GCC ACC GTG CAC CTG      624
Ala Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Thr Val His Leu
        195                 200                 205

GAG CAG GGC GGC ACC AAG CTG GTG AAG GAC TTA CGT CCC GGA GAC CGC      672
Glu Gln Gly Gly Thr Lys Leu Val Lys Asp Leu Arg Pro Gly Asp Arg
    210                 215                 220

GTG CTG GCG GCT GAC GAC CAG GGC CGG CTG CTG TAC AGC GAC TTC CTC      720
Val Leu Ala Ala Asp Asp Gln Gly Arg Leu Leu Tyr Ser Asp Phe Leu
225                 230                 235                 240

ACC TTC CTG GAC CGC GAC GAA GGC GCC AAG AAG GTC TTC TAC GTG ATC      768
Thr Phe Leu Asp Arg Asp Glu Gly Ala Lys Lys Val Phe Tyr Val Ile
                245                 250                 255

GAG ACG CTG GAG CCG CGC GAG CGC CTG CTG CTC ACC GCC GCG CAC CTG      816
Glu Thr Leu Glu Pro Arg Glu Arg Leu Leu Leu Thr Ala Ala His Leu
            260                 265                 270

CTC TTC GTG GCG CCG CAC AAC GAC TCG GGG CCC ACG CCC GGG CCA AGC      864
Leu Phe Val Ala Pro His Asn Asp Ser Gly Pro Thr Pro Gly Pro Ser
        275                 280                 285

GCG CTC TTT GCC AGC CGC GTG CGC CCC GGG CAG CGC GTG TAC GTG GTG      912
Ala Leu Phe Ala Ser Arg Val Arg Pro Gly Gln Arg Val Tyr Val Val
    290                 295                 300

GCT GAA CGC GGC GGG GAC CGC CGG CTG CTG CCC GCC GCG GTG CAC AGC      960
Ala Glu Arg Gly Gly Asp Arg Arg Leu Leu Pro Ala Ala Val His Ser
305                 310                 315                 320

GTG ACG CTG CGA GAG GAG GAG GCG GGC GCG TAC GCG CCG CTC ACG GCG     1008
Val Thr Leu Arg Glu Glu Glu Ala Gly Ala Tyr Ala Pro Leu Thr Ala
                325                 330                 335

CAC GGC ACC ATT CTC ATC AAC CGG GTG CTC GCC TCG TGC TAC GCT GTC     1056
His Gly Thr Ile Leu Ile Asn Arg Val Leu Ala Ser Cys Tyr Ala Val
            340                 345                 350

ATC GAG GAG CAC AGC TGG GCA CAC CGG GCC TTC GCG CCT TTC CGC CTG     1104
Ile Glu Glu His Ser Trp Ala His Arg Ala Phe Ala Pro Phe Arg Leu
        355                 360                 365

GCG CAC GCG CTG CTG GCC GCG CTG GCA CCC GCC CGC ACG GAC GGC GGG     1152
Ala His Ala Leu Leu Ala Ala Leu Ala Pro Ala Arg Thr Asp Gly Gly
    370                 375                 380

GGC GGG GGC AGC ATC CCT GCA GCG CAA TCT GCA ACG GAA GCG AGG GGC     1200
Gly Gly Gly Ser Ile Pro Ala Ala Gln Ser Ala Thr Glu Ala Arg Gly
385                 390                 395                 400

GCG GAG CCG ACT GCG GGC ATC CAC TGG TAC TCG CAG CTG CTC TAC CAC     1248
Ala Glu Pro Thr Ala Gly Ile His Trp Tyr Ser Gln Leu Leu Tyr His
                405                 410                 415
```

```
ATT GGC ACC TGG CTG TTG GAC AGC GAG ACC ATG CAT CCC TTG GGA ATG      1296
Ile Gly Thr Trp Leu Leu Asp Ser Glu Thr Met His Pro Leu Gly Met
            420                 425                 430

GCG GTC AAG TCC AGC TG                                               1313
Ala Val Lys Ser Ser
        435
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1256 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1257

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATG CGG CTT TTG ACG AGA GTG CTG CTG GTG TCT CTT CTC ACT CTG TCC        48
Met Arg Leu Leu Thr Arg Val Leu Leu Val Ser Leu Leu Thr Leu Ser
  1               5                  10                  15

TTG GTG GTG TCC GGA CTG GCC TGC GGT CCT GGC AGA GGC TAC GGC AGA        96
Leu Val Val Ser Gly Leu Ala Cys Gly Pro Gly Arg Gly Tyr Gly Arg
             20                  25                  30

AGA AGA CAT CCG AAG AAG CTG ACA CCT CTC GCC TAC AAG CAG TTC ATA       144
Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile
         35                  40                  45

CCT AAT GTC GCG GAG AAG ACC TTA GGG GCC AGC GGC AGA TAC GAG GGC       192
Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg Tyr Glu Gly
     50                  55                  60

AAG ATA ACG CGC AAT TCG GAG AGA TTT AAA GAA CTT ACT CCA AAT TAC       240
Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn Tyr
 65                  70                  75                  80

AAT CCC GAC ATT ATC TTT AAG GAT GAG GAG AAC ACG GGA GCG GAC AGG       288
Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg
                 85                  90                  95

CTC ATG ACA CAG AGA TGC AAA GAC AAG CTG AAC TCG CTG GCC ATC TCT       336
Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ser Leu Ala Ile Ser
            100                 105                 110

GTA ATG AAC CAC TGG CCA GGG GTT AAG CTG CGT GTG ACA GAG GGC TGG       384
Val Met Asn His Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp
        115                 120                 125

GAT GAG GAC GGT CAC CAT TTT GAA GAA TCA CTC CAC TAC GAG GGA AGA       432
Asp Glu Asp Gly His His Phe Glu Glu Ser Leu His Tyr Glu Gly Arg
```

```
    130                     135                     140

GCT GTT GAT ATT ACC ACC TCT GAC CGA GAC AAG AGC AAA TAC GGG ACA      480
Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Lys Ser Lys Tyr Gly Thr
145                 150                 155                 160

CTG TCT CGC CTA GCT GTG GAG GCT GGA TTT GAC TGG GTC TAT TAC GAG      528
Leu Ser Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu
                165                 170                 175

TCC AAA GCC CAC ATT CAT TGC TCT GTC AAA GCA GAA AAT TCG GTT GCT      576
Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val Ala
            180                 185                 190

GCG AAA TCT GGG GGC TGT TTC CCA GGT TCG GCT CTG GTC TCG CTC CAG      624
Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Leu Val Ser Leu Gln
        195                 200                 205

GAC GGA GGA CAG AAG GCC GTG AAG GAC CTG AAC CCC GGA GAC AAG GTG      672
Asp Gly Gly Gln Lys Ala Val Lys Asp Leu Asn Pro Gly Asp Lys Val
    210                 215                 220

CTG GCG GCA GAC AGC GCG GGA AAC CTG GTG TTC AGC GAC TTC ATC ATG      720
Leu Ala Ala Asp Ser Ala Gly Asn Leu Val Phe Ser Asp Phe Ile Met
225                 230                 235                 240

TTC ACA GAC CGA GAC TCC ACG ACG CGA CGT GTG TTT TAC GTC ATA GAA      768
Phe Thr Asp Arg Asp Ser Thr Thr Arg Arg Val Phe Tyr Val Ile Glu
                245                 250                 255

ACG CAA GAA CCC GTT GAA AAG ATC ACC CTC ACC GCC GCT CAC CTC CTT      816
Thr Gln Glu Pro Val Glu Lys Ile Thr Leu Thr Ala Ala His Leu Leu
            260                 265                 270

TTT GTC CTC GAC AAC TCA ACG GAA GAT CTC CAC ACC ATG ACC GCC GCG      864
Phe Val Leu Asp Asn Ser Thr Glu Asp Leu His Thr Met Thr Ala Ala
        275                 280                 285

TAT GCC AGC AGT GTC AGA GCC GGA CAA AAG GTG ATG GTT GTT GAT GAT      912
Tyr Ala Ser Ser Val Arg Ala Gly Gln Lys Val Met Val Val Asp Asp
    290                 295                 300

AGC GGT CAG CTT AAA TCT GTC ATC GTG CAG CGG ATA TAC ACG GAG GAG      960
Ser Gly Gln Leu Lys Ser Val Ile Val Gln Arg Ile Tyr Thr Glu Glu
305                 310                 315                 320

CAG CGG GGC TCG TTC GCA CCA GTG ACT GCA CAT GGG ACC ATT GTG GTC     1008
Gln Arg Gly Ser Phe Ala Pro Val Thr Ala His Gly Thr Ile Val Val
                325                 330                 335

GAC AGA ATA CTG GCG TCC TGT TAC GCC GTA ATA GAG GAC CAG GGG CTT     1056
Asp Arg Ile Leu Ala Ser Cys Tyr Ala Val Ile Glu Asp Gln Gly Leu
            340                 345                 350

GCG CAT TTG GCC TTC GCG CCC GCC AGG CTC TAT TAT TAC GTG TCA TCA     1104
Ala His Leu Ala Phe Ala Pro Ala Arg Leu Tyr Tyr Tyr Val Ser Ser
        355                 360                 365

TTC CTG TCC CCC AAA ACT CCA GCA GTC GGT CCA ATG CGA CTT TAC AAC     1152
Phe Leu Ser Pro Lys Thr Pro Ala Val Gly Pro Met Arg Leu Tyr Asn
    370                 375                 380
```

```
AGG AGG GGG TCC ACT GGT ACT CCA GGC TCC TGT CAT CAA ATG GGA ACG      1200
Arg Arg Gly Ser Thr Gly Thr Pro Gly Ser Cys His Gln Met Gly Thr
385                 390                 395                 400

TGG CTT TTG GAC AGC AAC ATG CTT CAT CCT TTG GGG ATG TCA GTA AAC      1248
Trp Leu Leu Asp Ser Asn Met Leu His Pro Leu Gly Met Ser Val Asn
                405                 410                 415

TCA AGC TG                                                           1256
Ser Ser
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1425 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1425

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
ATG CTG CTG CTG GCG AGA TGT CTG CTG CTA GTC CTC GTC TCC TCG CTG        48
Met Leu Leu Leu Ala Arg Cys Leu Leu Leu Val Leu Val Ser Ser Leu
  1               5                  10                  15

CTG GTA TGC TCG GGA CTG GCG TGC GGA CCG GGC AGG GGG TTC GGG AAG        96
Leu Val Cys Ser Gly Leu Ala Cys Gly Pro Gly Arg Gly Phe Gly Lys
             20                  25                  30

AGG AGG CAC CCC AAA AAG CTG ACC CCT TTA GCC TAC AAG CAG TTT ATC       144
Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile
         35                  40                  45

CCC AAT GTG GCC GAG AAG ACC CTA GGC GCC AGC GGA AGG TAT GAA GGG       192
Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg Tyr Glu Gly
     50                  55                  60

AAG ATC TCC AGA AAC TCC GAG CGA TTT AAG GAA CTC ACC CCC AAT TAC       240
Lys Ile Ser Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn Tyr
 65                  70                  75                  80

AAC CCC GAC ATC ATA TTT AAG GAT GAA GAA AAC ACC GGA GCG GAC AGG       288
Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg
                 85                  90                  95

CTG ATG ACT CAG AGG TGT AAG GAC AAG TTG AAC GCT TTG GCC ATC TCG       336
Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala Ile Ser
            100                 105                 110

GTG ATG AAC CAG TGG CCA GGA GTG AAA CTG CGG GTG ACC GAG GGC TGG       384
Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp
        115                 120                 125
```

```
GAC GAA GAT GGC CAC CAC TCA GAG GAG TCT CTG CAC TAC GAG GGC CGC      432
Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr Glu Gly Arg
    130                 135                 140

GCA GTG GAC ATC ACC ACG TCT GAC CGC GAC CGC AGC AAG TAC GGC ATG      480
Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly Met
145                 150                 155                 160

CTG GCC CGC CTG GCG GTG GAG GCC GGC TTC GAC TGG GTG TAC TAC GAG      528
Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu
                165                 170                 175

TCC AAG GCA CAT ATC CAC TGC TCG GTG AAA GCA GAG AAC TCG GTG GCG      576
Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val Ala
            180                 185                 190

GCC AAA TCG GGA GGC TGC TTC CCG GGC TCG GCC ACG GTG CAC CTG GAG      624
Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Thr Val His Leu Glu
        195                 200                 205

CAG GGC GGC ACC AAG CTG GTG AAG GAC CTG AGC CCC GGG GAC CGC GTG      672
Gln Gly Gly Thr Lys Leu Val Lys Asp Leu Ser Pro Gly Asp Arg Val
    210                 215                 220

CTG GCG GCG GAC GAC CAG GGC CGG CTG CTC TAC AGC GAC TTC CTC ACT      720
Leu Ala Ala Asp Asp Gln Gly Arg Leu Leu Tyr Ser Asp Phe Leu Thr
225                 230                 235                 240

TTC CTG GAC CGC GAC GAC GGC GCC AAG AAG GTC TTC TAC GTG ATC GAG      768
Phe Leu Asp Arg Asp Asp Gly Ala Lys Lys Val Phe Tyr Val Ile Glu
                245                 250                 255

ACG CGG GAG CCG CGC GAG CGC CTG CTG CTC ACC GCC GCG CAC CTG CTC      816
Thr Arg Glu Pro Arg Glu Arg Leu Leu Leu Thr Ala Ala His Leu Leu
            260                 265                 270

TTT GTG GCG CCG CAC AAC GAC TCG GCC ACC GGG GAG CCC GAG GCG TCC      864
Phe Val Ala Pro His Asn Asp Ser Ala Thr Gly Glu Pro Glu Ala Ser
        275                 280                 285

TCG GGC TCG GGG CCG CCT TCC GGG GGC GCA CTG GGG CCT CGG GCG CTG      912
Ser Gly Ser Gly Pro Pro Ser Gly Gly Ala Leu Gly Pro Arg Ala Leu
    290                 295                 300

TTC GCC AGC CGC GTG CGC CCG GGC CAG CGC GTG TAC GTG GTG GCC GAG      960
Phe Ala Ser Arg Val Arg Pro Gly Gln Arg Val Tyr Val Val Ala Glu
305                 310                 315                 320

CGT GAC GGG GAC CGC CGG CTC CTG CCC GCC GCT GTG CAC AGC GTG ACC     1008
Arg Asp Gly Asp Arg Arg Leu Leu Pro Ala Ala Val His Ser Val Thr
                325                 330                 335

CTA AGC GAG GAG GCC GCG GGC GCC TAC GCG CCG CTC ACG GCC CAG GGC     1056
Leu Ser Glu Glu Ala Ala Gly Ala Tyr Ala Pro Leu Thr Ala Gln Gly
            340                 345                 350

ACC ATT CTC ATC AAC CGG GTG CTG GCC TCG TGC TAC GCG GTC ATC GAG     1104
Thr Ile Leu Ile Asn Arg Val Leu Ala Ser Cys Tyr Ala Val Ile Glu
        355                 360                 365

GAG CAC AGC TGG GCG CAC CGG GCC TTC GCG CCC TTC CGC CTG GCG CAC     1152
```

Glu His Ser Trp Ala His Arg Ala Phe Ala Pro Phe Arg Leu Ala His
370 375 380

GCG CTC CTG GCT GCA CTG GCG CCC GCG CGC ACG GAC CGC GGC GGG GAC 1200
Ala Leu Leu Ala Ala Leu Ala Pro Ala Arg Thr Asp Arg Gly Gly Asp
385 390 395 400

AGC GGC GGC GGG GAC CGC GGG GGC GGC GGC GGC AGA GTA GCC CTA ACC 1248
Ser Gly Gly Gly Asp Arg Gly Gly Gly Gly Gly Arg Val Ala Leu Thr
405 410 415

GCT CCA GGT GCT GCC GAC GCT CCG GGT GCG GGG GCC ACC GCG GGC ATC 1296
Ala Pro Gly Ala Ala Asp Ala Pro Gly Ala Gly Ala Thr Ala Gly Ile
420 425 430

CAC TGG TAC TCG CAG CTG CTC TAC CAA ATA GGC ACC TGG CTC CTG GAC 1344
His Trp Tyr Ser Gln Leu Leu Tyr Gln Ile Gly Thr Trp Leu Leu Asp
435 440 445

AGC GAG GCC CTG CAC CCG CTG GGC ATG GCG GTC AAG TCC AGC NNN AGC 1392
Ser Glu Ala Leu His Pro Leu Gly Met Ala Val Lys Ser Ser Xaa Ser
450 455 460

CGG GGG GCC GGG GGA GGG GCG CGG GAG GGG GCC 1425
Arg Gly Ala Gly Gly Gly Ala Arg Glu Gly Ala
465 470 475

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 51..1283

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CATCAGCCCA CCAGGAGACC TCGCCCGCCG CTCCCCCGGG CTCCCCGGCC ATG TCT 56
Met Ser
1

CCC GCC CGG CTC CGG CCC CGA CTG CAC TTC TGC CTG GTC CTG TTG CTG 104
Pro Ala Arg Leu Arg Pro Arg Leu His Phe Cys Leu Val Leu Leu Leu
5 10 15

CTG CTG GTG GTG CCC GCG GCA TGG GGC TGC GGG CCG GGT CGG GTG GTG 152
Leu Leu Val Val Pro Ala Ala Trp Gly Cys Gly Pro Gly Arg Val Val
20 25 30

GGC AGC CGC CGG CGA CCG CCA CGC AAA CTC GTG CCG CTC GCC TAC AAG 200
Gly Ser Arg Arg Arg Pro Pro Arg Lys Leu Val Pro Leu Ala Tyr Lys
35 40 45 50

```
CAG TTC AGC CCC AAT GTG CCC GAG AAG ACC CTG GGC GCC AGC GGA CGC     248
Gln Phe Ser Pro Asn Val Pro Glu Lys Thr Leu Gly Ala Ser Gly Arg
                 55                  60                  65

TAT GAA GGC AAG ATC GCT CGC AGC TCC GAG CGC TTC AAG GAG CTC ACC     296
Tyr Glu Gly Lys Ile Ala Arg Ser Ser Glu Arg Phe Lys Glu Leu Thr
             70                  75                  80

CCC AAT TAC AAT CCA GAC ATC ATC TTC AAG GAC GAG GAG AAC ACA GGC     344
Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly
         85                  90                  95

GCC GAC CGC CTC ATG ACC CAG CGC TGC AAG GAC CGC CTG AAC TCG CTG     392
Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Arg Leu Asn Ser Leu
    100                 105                 110

GCT ATC TCG GTG ATG AAC CAG TGG CCC GGT GTG AAG CTG CGG GTG ACC     440
Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr
115                 120                 125                 130

GAG GGC TGG GAC GAG GAC GGC CAC CAC TCA GAG GAG TCC CTG CAT TAT     488
Glu Gly Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr
                135                 140                 145

GAG GGC CGC GCG GTG GAC ATC ACC ACA TCA GAC CGC GAC CGC AAT AAG     536
Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys
            150                 155                 160

TAT GGA CTG CTG GCG CGC TTG GCA GTG GAG GCC GGC TTT GAC TGG GTG     584
Tyr Gly Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val
        165                 170                 175

TAT TAC GAG TCA AAG GCC CAC GTG CAT TGC TCC GTC AAG TCC GAG CAC     632
Tyr Tyr Glu Ser Lys Ala His Val His Cys Ser Val Lys Ser Glu His
    180                 185                 190

TCG GCC GCA GCC AAG ACG GGC GGC TGC TTC CCT GCC GGA GCC CAG GTA     680
Ser Ala Ala Ala Lys Thr Gly Gly Cys Phe Pro Ala Gly Ala Gln Val
195                 200                 205                 210

CGC CTG GAG AGT GGG GCG CGT GTG GCC TTG TCA GCC GTG AGG CCG GGA     728
Arg Leu Glu Ser Gly Ala Arg Val Ala Leu Ser Ala Val Arg Pro Gly
                215                 220                 225

GAC CGT GTG CTG GCC ATG GGG GAG GAT GGG AGC CCC ACC TTC AGC GAT     776
Asp Arg Val Leu Ala Met Gly Glu Asp Gly Ser Pro Thr Phe Ser Asp
            230                 235                 240

GTG CTC ATT TTC CTG GAC CGC GAG CCC CAC AGG CTG AGA GCC TTC CAG     824
Val Leu Ile Phe Leu Asp Arg Glu Pro His Arg Leu Arg Ala Phe Gln
        245                 250                 255

GTC ATC GAG ACT CAG GAC CCC CCA CGC CGC CTG GCA CTC ACA CCC GCT     872
Val Ile Glu Thr Gln Asp Pro Pro Arg Arg Leu Ala Leu Thr Pro Ala
    260                 265                 270

CAC CTG CTC TTT ACG GCT GAC AAT CAC ACG GAG CCG GCA GCC CGC TTC     920
His Leu Leu Phe Thr Ala Asp Asn His Thr Glu Pro Ala Ala Arg Phe
275                 280                 285                 290

CGG GCC ACA TTT GCC AGC CAC GTG CAG CCT GGC CAG TAC GTG CTG GTG     968
```

```
Arg Ala Thr Phe Ala Ser His Val Gln Pro Gly Gln Tyr Val Leu Val
                295                 300                 305

GCT GGG GTG CCA GGC CTG CAG CCT GCC CGC GTG GCA GCT GTC TCT ACA     1016
Ala Gly Val Pro Gly Leu Gln Pro Ala Arg Val Ala Ala Val Ser Thr
            310                 315                 320

CAC GTG GCC CTC GGG GCC TAC GCC CCG CTC ACA AAG CAT GGG ACA CTG     1064
His Val Ala Leu Gly Ala Tyr Ala Pro Leu Thr Lys His Gly Thr Leu
        325                 330                 335

GTG GTG GAG GAT GTG GTG GCA TCC TGC TTC GCG GCC GTG GCT GAC CAC     1112
Val Val Glu Asp Val Val Ala Ser Cys Phe Ala Ala Val Ala Asp His
    340                 345                 350

CAC CTG GCT CAG TTG GCC TTC TGG CCC CTG AGA CTC TTT CAC AGC TTG     1160
His Leu Ala Gln Leu Ala Phe Trp Pro Leu Arg Leu Phe His Ser Leu
355                 360                 365                 370

GCA TGG GGC AGC TGG ACC CCG GGG GAG GGT GTG CAT TGG TAC CCC CAG     1208
Ala Trp Gly Ser Trp Thr Pro Gly Glu Gly Val His Trp Tyr Pro Gln
                375                 380                 385

CTG CTC TAC CGC CTG GGG CGT CTC CTG CTA GAA GAG GGC AGC TTC CAC     1256
Leu Leu Tyr Arg Leu Gly Arg Leu Leu Leu Glu Glu Gly Ser Phe His
            390                 395                 400

CCA CTG GGC ATG TCC GGG GCA GGG AGC TGAAAGGACT CCACCGCTGC           1303
Pro Leu Gly Met Ser Gly Ala Gly Ser
        405                 410

CCTCCTGGAA CTGCTGTACT GGGTCCAGAA GCCTCTCAGC CAGGAGGGAG CTGGCCCTGG  1363

AAGGGACCTG AGCTGGGGGA CACTGGCTCC TGCCATCTCC TCTGCCATGA AGATACACCA  1423

TTGAGACTTG ACTGGGCAAC ACCAGCGTCC CCCACCCGCG TCGTGGTGTA GTCATAGAGC  1483

TGCAAGCTGA GCTGGCGAGG GGATGGTTGT TGACCCCTCT CTCCTAGAGA CCTTGAGGCT  1543

GGCACGGCGA CTCCCAACTC AGCCTGCTCT CACTACGAGT TTTCATACTC TGCCTCCCCC  1603

ATTGGGAGGG CCCATTCCC                                                1622
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1191 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1191

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
ATG GCT CTC CTG ACC AAT CTA CTG CCC TTG TGC TGC TTG GCA CTT CTG       48
```

```
Met Ala Leu Leu Thr Asn Leu Leu Pro Leu Cys Cys Leu Ala Leu Leu
 1               5                  10                  15

GCG CTG CCA GCC CAG AGC TGC GGG CCG GGC CGG GGG CCG GTT GGC CGG       96
Ala Leu Pro Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val Gly Arg
            20                  25                  30

CGC CGC TAT GCG CGC AAG CAG CTC GTG CCG CTA CTC TAC AAG CAA TTT      144
Arg Arg Tyr Ala Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe
        35                  40                  45

GTG CCC GGC GTG CCA GAG CGG ACC CTG GGC GCC AGT GGG CCA GCG GAG      192
Val Pro Gly Val Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu
    50                  55                  60

GGG AGG GTG GCA AGG GGC TCC GAG CGC TTC CGG GAC CTC GTG CCC AAC      240
Gly Arg Val Ala Arg Gly Ser Glu Arg Phe Arg Asp Leu Val Pro Asn
65                  70                  75                  80

TAC AAC CCC GAC ATC ATC TTC AAG GAT GAG GAG AAC AGT GGA GCC GAC      288
Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp
                85                  90                  95

CGC CTG ATG ACC GAG CGT TGC AAG GAG AGG GTG AAC GCT TTG GCC ATT      336
Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile
            100                 105                 110

GCC GTG ATG AAC ATG TGG CCC GGA GTG CGC CTA CGA GTG ACT GAG GGC      384
Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly
        115                 120                 125

TGG GAC GAG GAC GGC CAC CAC GCT CAG GAT TCA CTC CAC TAC GAA GGC      432
Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly
    130                 135                 140

CGT GCT TTG GAC ATC ACT ACG TCT GAC CGC GAC CGC AAC AAG TAT GGG      480
Arg Ala Leu Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly
145                 150                 155                 160

TTG CTG GCG CGC CTC GCA GTG GAA GCC GGC TTC GAC TGG GTC TAC TAC      528
Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
                165                 170                 175

GAG TCC CGC AAC CAC GTC CAC GTG TCG GTC AAA GCT GAT AAC TCA CTG      576
Glu Ser Arg Asn His Val His Val Ser Val Lys Ala Asp Asn Ser Leu
            180                 185                 190

GCG GTC CGG GCG GGC GGC TGC TTT CCG GGA AAT GCA ACT GTG CGC CTG      624
Ala Val Arg Ala Gly Gly Cys Phe Pro Gly Asn Ala Thr Val Arg Leu
        195                 200                 205

TGG AGC GGC GAG CGG AAA GGG CTG CGG GAA CTG CAC CGC GGA GAC TGG      672
Trp Ser Gly Glu Arg Lys Gly Leu Arg Glu Leu His Arg Gly Asp Trp
    210                 215                 220

GTT TTG GCG GCC GAT GCG TCA GGC CGG GTG GTG CCC ACG CCG GTG CTG      720
Val Leu Ala Ala Asp Ala Ser Gly Arg Val Val Pro Thr Pro Val Leu
225                 230                 235                 240

CTC TTC CTG GAC CGG GAC TTG CAG CGC CGG GCT TCA TTT GTG GCT GTG      768
Leu Phe Leu Asp Arg Asp Leu Gln Arg Arg Ala Ser Phe Val Ala Val
                245                 250                 255
```

```
GAG ACC GAG TGG CCT CCA CGC AAA CTG TTG CTC ACG CCC TGG CAC CTG      816
Glu Thr Glu Trp Pro Pro Arg Lys Leu Leu Leu Thr Pro Trp His Leu
            260                 265                 270

GTG TTT GCC GCT CGA GGG CCG GCG CCC GCG CCA GGC GAC TTT GCA CCG      864
Val Phe Ala Ala Arg Gly Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro
        275                 280                 285

GTG TTC GCG CGC CGG CTA CGC GCT GGG GAC TCG GTG CTG GCG CCC GGC      912
Val Phe Ala Arg Arg Leu Arg Ala Gly Asp Ser Val Leu Ala Pro Gly
    290                 295                 300

GGG GAT GCG CTT CGG CCA GCG CGC GTG GCC CGT GTG GCG CGG GAG GAA      960
Gly Asp Ala Leu Arg Pro Ala Arg Val Ala Arg Val Ala Arg Glu Glu
305                 310                 315                 320

GCC GTG GGC GTG TTC GCG CCG CTC ACC GCG CAC GGG ACG CTG CTG GTG     1008
Ala Val Gly Val Phe Ala Pro Leu Thr Ala His Gly Thr Leu Leu Val
                325                 330                 335

AAC GAT GTC CTG GCC TCT TGC TAC GCG GTT CTG GAG AGT CAC CAG TGG     1056
Asn Asp Val Leu Ala Ser Cys Tyr Ala Val Leu Glu Ser His Gln Trp
            340                 345                 350

GCG CAC CGC GCT TTT GCC CCC TTG AGA CTG CTG CAC GCG CTA GGG GCG     1104
Ala His Arg Ala Phe Ala Pro Leu Arg Leu Leu His Ala Leu Gly Ala
        355                 360                 365

CTG CTC CCC GGC GGG GCC GTC CAG CCG ACT GGC ATG CAT TGG TAC TCT     1152
Leu Leu Pro Gly Gly Ala Val Gln Pro Thr Gly Met His Trp Tyr Ser
    370                 375                 380

CGG CTC CTC TAC CGC TTA GCG GAG GAG CTA CTG GGC TG                  1191
Arg Leu Leu Tyr Arg Leu Ala Glu Glu Leu Leu Gly
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1251 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1248

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATG GAC GTA AGG CTG CAT CTG AAG CAA TTT GCT TTA CTG TGT TTT ATC       48
Met Asp Val Arg Leu His Leu Lys Gln Phe Ala Leu Leu Cys Phe Ile
 1               5                  10                  15

AGC TTG CTT CTG ACG CCT TGT GGA TTA GCC TGT GGT CCT GGT AGA GGT       96
```

```
Ser Leu Leu Leu Thr Pro Cys Gly Leu Ala Cys Gly Pro Gly Arg Gly
            20                  25                  30

TAT GGA AAA CGA AGA CAC CCA AAG AAA TTA ACC CCG TTG GCT TAC AAG      144
Tyr Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys
        35                  40                  45

CAA TTC ATC CCC AAC GTT GCT GAG AAA ACG CTT GGA GCC AGC GGC AAA      192
Gln Phe Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Lys
    50                  55                  60

TAC GAA GGC AAA ATC ACA AGG AAT TCA GAG AGA TTT AAA GAG CTG ATT      240
Tyr Glu Gly Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Ile
65                  70                  75                  80

CCG AAT TAT AAT CCC GAT ATC ATC TTT AAG GAC GAG GAA AAC ACA AAC      288
Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Asn
                85                  90                  95

GCT GAC AGG CTG ATG ACC AAG CGC TGT AAG GAC AAG TTA AAT TCG TTG      336
Ala Asp Arg Leu Met Thr Lys Arg Cys Lys Asp Lys Leu Asn Ser Leu
            100                 105                 110

GCC ATA TCC GTC ATG AAC CAC TGG CCC GGC GTG AAA CTG CGC GTC ACT      384
Ala Ile Ser Val Met Asn His Trp Pro Gly Val Lys Leu Arg Val Thr
        115                 120                 125

GAA GGC TGG GAT GAG GAT GGT CAC CAT TTA GAA GAA TCT TTG CAC TAT      432
Glu Gly Trp Asp Glu Asp Gly His His Leu Glu Glu Ser Leu His Tyr
    130                 135                 140

GAG GGA CGG GCA GTG GAC ATC ACT ACC TCA GAC AGG GAT AAA AGC AAG      480
Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Lys Ser Lys
145                 150                 155                 160

TAT  GGG  ATG  CTA  TCC  AGG  CTT  GCA  GTG  GAG  GCA  GGA  TTC  GAC  TGG  GTC
528
Tyr Gly Met Leu Ser Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val
                165                 170                 175

TAT  TAT  GAA  TCT  AAA  GCC  CAC  ATA  CAC  TGC  TCT  GTC  AAA  GCA  GAA  AAT
576
Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn
            180                 185                 190

TCA  GTG  GCT  GCT  AAA  TCA  GGA  GGA  TGT  TTT  CCT  GGG  TCT  GGG  ACG  GTG
624
Ser Val Ala Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Gly Thr Val
        195                 200                 205

ACA  CTT  GGT  GAT  GGG  ACG  AGG  AAA  CCC  ATC  AAA  GAT  CTT  AAA  GTG  GGC
672
Thr Leu Gly Asp Gly Thr Arg Lys Pro Ile Lys Asp Leu Lys Val Gly
    210                 215                 220

GAC  CGG  GTT  TTG  GCT  GCA  GAC  GAG  AAG  GGA  AAT  GTC  TTA  ATA  AGC  GAC
720
Asp Arg Val Leu Ala Ala Asp Glu Lys Gly Asn Val Leu Ile Ser Asp
225                 230                 235                 240

TTT  ATT  ATG  TTT  ATA  GAC  CAC  GAT  CCG  ACA  ACG  AGA  AGG  CAA  TTC  ATC
768
```

```
Phe Ile Met Phe Ile Asp His Asp Pro Thr Thr Arg Arg Gln Phe Ile
                245                 250                 255

GTC ATC GAG ACG TCA GAA CCT TTC ACC AAG CTC ACC CTC ACT GCC GCG
816
Val Ile Glu Thr Ser Glu Pro Phe Thr Lys Leu Thr Leu Thr Ala Ala
            260                 265                 270

CAC CTA GTT TTC GTT GGA AAC TCT TCA GCA GCT TCG GGT ATA ACA GCA
864
His Leu Val Phe Val Gly Asn Ser Ser Ala Ala Ser Gly Ile Thr Ala
        275                 280                 285

ACA TTT GCC AGC AAC GTG AAG CCT GGA GAT ACA GTT TTA GTG TGG GAA
912
Thr Phe Ala Ser Asn Val Lys Pro Gly Asp Thr Val Leu Val Trp Glu
    290                 295                 300

GAC ACA TGC GAG AGC CTC AAG AGC GTT ACA GTG AAA AGG ATT TAC ACT
960
Asp Thr Cys Glu Ser Leu Lys Ser Val Thr Val Lys Arg Ile Tyr Thr
305                 310                 315                 320

GAG GAG CAC GAG GGC TCT TTT GCG CCA GTC ACC GCG CAC GGA ACC ATA
1008
Glu Glu His Glu Gly Ser Phe Ala Pro Val Thr Ala His Gly Thr Ile
                325                 330                 335

ATA GTG GAT CAG GTG TTG GCA TCG TGC TAC GCG GTC ATT GAG AAC CAC
1056
Ile Val Asp Gln Val Leu Ala Ser Cys Tyr Ala Val Ile Glu Asn His
            340                 345                 350

AAA TGG GCA CAT TGG GCT TTT GCG CCG GTC AGG TTG TGT CAC AAG CTG
1104
Lys Trp Ala His Trp Ala Phe Ala Pro Val Arg Leu Cys His Lys Leu
        355                 360                 365

ATG ACG TGG CTT TTT CCG GCT CGT GAA TCA AAC GTC AAT TTT CAG GAG
1152
Met Thr Trp Leu Phe Pro Ala Arg Glu Ser Asn Val Asn Phe Gln Glu
    370                 375                 380

GAT GGT ATC CAC TGG TAC TCA AAT ATG CTG TTT CAC ATC GGC TCT TGG
1200
Asp Gly Ile His Trp Tyr Ser Asn Met Leu Phe His Ile Gly Ser Trp
385                 390                 395                 400

CTG CTG GAC AGA GAC TCT TTC CAT CCA CTC GGG ATT TTA CAC TTA AGT
1248
Leu Leu Asp Arg Asp Ser Phe His Pro Leu Gly Ile Leu His Leu Ser
                405                 410                 415

TGA
1251
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 425 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Met Val Glu Met Leu Leu Leu Thr Arg Ile Leu Leu Val Gly Phe Ile
 1               5                  10                  15

Cys Ala Leu Leu Val Ser Ser Gly Leu Thr Cys Gly Pro Gly Arg Gly
            20                  25                  30

Ile Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys
        35                  40                  45

Gln Phe Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg
    50                  55                  60

Tyr Glu Gly Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr
65                  70                  75                  80

Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly
                85                  90                  95

Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu
            100                 105                 110

Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr
        115                 120                 125

Glu Gly Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr
    130                 135                 140

Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys
145                 150                 155                 160

Tyr Gly Met Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val
                165                 170                 175

Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn
            180                 185                 190

Ser Val Ala Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Thr Val
        195                 200                 205

His Leu Glu His Gly Gly Thr Lys Leu Val Lys Asp Leu Ser Pro Gly
    210                 215                 220

Asp Arg Val Leu Ala Ala Asp Ala Asp Gly Arg Leu Leu Tyr Ser Asp
225                 230                 235                 240

Phe Leu Thr Phe Leu Asp Arg Met Asp Ser Ser Arg Lys Leu Phe Tyr
                245                 250                 255

Val Ile Glu Thr Arg Gln Pro Arg Ala Arg Leu Leu Leu Thr Ala Ala
            260                 265                 270

His Leu Leu Phe Val Ala Pro Gln His Asn Gln Ser Glu Ala Thr Gly
        275                 280                 285

Ser Thr Ser Gly Gln Ala Leu Phe Ala Ser Asn Val Lys Pro Gly Gln
```

290 295 300

Arg Val Tyr Val Leu Gly Glu Gly Gly Gln Gln Leu Leu Pro Ala Ser
305 310 315 320

Val His Ser Val Ser Leu Arg Glu Glu Ala Ser Gly Ala Tyr Ala Pro
325 330 335

Leu Thr Ala Gln Gly Thr Ile Leu Ile Asn Arg Val Leu Ala Ser Cys
340 345 350

Tyr Ala Val Ile Glu Glu His Ser Trp Ala His Trp Ala Phe Ala Pro
355 360 365

Phe Arg Leu Ala Gln Gly Leu Leu Ala Ala Leu Cys Pro Asp Gly Ala
370 375 380

Ile Pro Thr Ala Ala Thr Thr Thr Thr Gly Ile His Trp Tyr Ser Arg
385 390 395 400

Leu Leu Tyr Arg Ile Gly Ser Trp Val Leu Asp Gly Asp Ala Leu His
405 410 415

Pro Leu Gly Met Val Ala Pro Ala Ser
420 425

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Ala Leu Pro Ala Ser Leu Leu Pro Leu Cys Cys Leu Ala Leu Leu
1 5 10 15

Ala Leu Ser Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val Gly Arg
20 25 30

Arg Arg Tyr Val Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe
35 40 45

Val Pro Ser Met Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu
50 55 60

Gly Arg Val Thr Arg Gly Ser Glu Arg Phe Arg Asp Leu Val Pro Asn
65 70 75 80

Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp
85 90 95

Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile
100 105 110

Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly

```
        115                 120                 125

Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly
    130                 135                 140

Arg Ala Leu Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly
145                 150                 155                     160

Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
                165                 170                 175

Glu Ser Arg Asn His Ile His Val Ser Val Lys Ala Asp Asn Ser Leu
            180                 185                 190

Ala Val Arg Ala Gly Gly Cys Phe Pro Gly Asn Ala Thr Val Arg Leu
        195                 200                 205

Arg Ser Gly Glu Arg Lys Gly Leu Arg Glu Leu His Arg Gly Asp Trp
    210                 215                 220

Val Leu Ala Ala Asp Ala Ala Gly Arg Val Val Pro Thr Pro Val Leu
225                 230                 235                     240

Leu Phe Leu Asp Arg Asp Leu Gln Arg Arg Ala Ser Phe Val Ala Val
                245                 250                 255

Glu Thr Glu Arg Pro Pro Arg Lys Leu Leu Leu Thr Pro Trp His Leu
            260                 265                 270

Val Phe Ala Ala Arg Gly Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro
        275                 280                 285

Val Phe Ala Arg Arg Leu Arg Ala Gly Asp Ser Val Leu Ala Pro Gly
    290                 295                 300

Gly Asp Ala Leu Gln Pro Ala Arg Val Ala Arg Val Ala Arg Glu Glu
305                 310                 315                     320

Ala Val Gly Val Phe Ala Pro Leu Thr Ala His Gly Thr Leu Leu Val
                325                 330                 335

Asn Asp Val Leu Ala Ser Cys Tyr Ala Val Leu Glu Ser His Gln Trp
            340                 345                 350

Ala His Arg Ala Phe Ala Pro Leu Arg Leu Leu His Ala Leu Gly Ala
        355                 360                 365

Leu Leu Pro Gly Gly Ala Val Gln Pro Thr Gly Met His Trp Tyr Ser
    370                 375                 380

Arg Leu Leu Tyr Arg Leu Ala Glu Glu Leu Met Gly
385                 390                 395
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 411 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Met Ser Pro Ala Trp Leu Arg Pro Arg Leu Arg Phe Cys Leu Phe Leu
 1               5                  10                  15

Leu Leu Leu Leu Leu Val Pro Ala Ala Arg Gly Cys Gly Pro Gly Arg
            20                  25                  30

Val Val Gly Ser Arg Arg Arg Pro Pro Arg Lys Leu Val Pro Leu Ala
        35                  40                  45

Tyr Lys Gln Phe Ser Pro Asn Val Pro Glu Lys Thr Leu Gly Ala Ser
    50                  55                  60

Gly Arg Tyr Glu Gly Lys Ile Ala Arg Ser Ser Glu Arg Phe Lys Glu
65                  70                  75                  80

Leu Thr Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn
                85                  90                  95

Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Arg Leu Asn
            100                 105                 110

Ser Leu Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg
        115                 120                 125

Val Thr Glu Gly Arg Asp Glu Asp Gly His His Ser Glu Glu Ser Leu
    130                 135                 140

His Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg
145                 150                 155                 160

Asn Lys Tyr Gly Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp
                165                 170                 175

Trp Val Tyr Tyr Glu Ser Lys Ala His Val His Cys Ser Val Lys Ser
            180                 185                 190

Glu His Ser Ala Ala Ala Lys Thr Gly Gly Cys Phe Pro Ala Gly Ala
        195                 200                 205

Gln Val Arg Leu Glu Asn Gly Glu Arg Val Ala Leu Ser Ala Val Lys
    210                 215                 220

Pro Gly Asp Arg Val Leu Ala Met Gly Glu Asp Gly Thr Pro Thr Phe
225                 230                 235                 240

Ser Asp Val Leu Ile Phe Leu Asp Arg Glu Pro Asn Arg Leu Arg Ala
                245                 250                 255

Phe Gln Val Ile Glu Thr Gln Asp Pro Pro Arg Arg Leu Ala Leu Thr
            260                 265                 270

Pro Ala His Leu Leu Phe Ile Ala Asp Asn His Thr Glu Pro Ala Ala
        275                 280                 285

His Phe Arg Ala Thr Phe Ala Ser His Val Gln Pro Gly Gln Tyr Val
    290                 295                 300
```

```
Leu Val Ser Gly Val Pro Gly Leu Gln Pro Ala Arg Val Ala Ala Val
305                 310                 315                 320

Ser Thr His Val Ala Leu Gly Ser Tyr Ala Pro Leu Thr Arg His Gly
                325                 330                 335

Thr Leu Val Val Glu Asp Val Val Ala Ser Cys Phe Ala Ala Val Ala
            340                 345                 350

Asp His His Leu Ala Gln Leu Ala Phe Trp Pro Leu Arg Leu Phe Pro
        355                 360                 365

Ser Leu Ala Trp Gly Ser Trp Thr Pro Ser Glu Gly Val His Ser Tyr
    370                 375                 380

Pro Gln Met Leu Tyr Arg Leu Gly Arg Leu Leu Leu Glu Glu Ser Thr
385                 390                 395                 400

Phe His Pro Leu Gly Met Ser Gly Ala Gly Ser
                405                 410
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 437 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Met Leu Leu Leu Leu Ala Arg Cys Phe Leu Val Ile Leu Ala Ser Ser
  1               5                  10                  15

Leu Leu Val Cys Pro Gly Leu Ala Cys Gly Pro Gly Arg Gly Phe Gly
             20                  25                  30

Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe
         35                  40                  45

Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg Tyr Glu
     50                  55                  60

Gly Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn
 65                  70                  75                  80

Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp
                 85                  90                  95

Arg Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala Ile
            100                 105                 110

Ser Val Met Asn Gln Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly
        115                 120                 125

Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr Glu Gly
    130                 135                 140
```

```
Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly
145                 150                 155                 160

Met Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
                165                 170                 175

Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val
            180                 185                 190

Ala Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Thr Val His Leu
        195                 200                 205

Glu Gln Gly Gly Thr Lys Leu Val Lys Asp Leu Arg Pro Gly Asp Arg
    210                 215                 220

Val Leu Ala Ala Asp Asp Gln Gly Arg Leu Leu Tyr Ser Asp Phe Leu
225                 230                 235                 240

Thr Phe Leu Asp Arg Asp Glu Gly Ala Lys Lys Val Phe Tyr Val Ile
                245                 250                 255

Glu Thr Leu Glu Pro Arg Glu Arg Leu Leu Leu Thr Ala Ala His Leu
            260                 265                 270

Leu Phe Val Ala Pro His Asn Asp Ser Gly Pro Thr Pro Gly Pro Ser
        275                 280                 285

Ala Leu Phe Ala Ser Arg Val Arg Pro Gly Gln Arg Val Tyr Val Val
    290                 295                 300

Ala Glu Arg Gly Gly Asp Arg Arg Leu Leu Pro Ala Ala Val His Ser
305                 310                 315                 320

Val Thr Leu Arg Glu Glu Glu Ala Gly Ala Tyr Ala Pro Leu Thr Ala
                325                 330                 335

His Gly Thr Ile Leu Ile Asn Arg Val Leu Ala Ser Cys Tyr Ala Val
            340                 345                 350

Ile Glu Glu His Ser Trp Ala His Arg Ala Phe Ala Pro Phe Arg Leu
        355                 360                 365

Ala His Ala Leu Leu Ala Ala Leu Ala Pro Ala Arg Thr Asp Gly Gly
    370                 375                 380

Gly Gly Gly Ser Ile Pro Ala Ala Gln Ser Ala Thr Glu Ala Arg Gly
385                 390                 395                 400

Ala Glu Pro Thr Ala Gly Ile His Trp Tyr Ser Gln Leu Leu Tyr His
                405                 410                 415

Ile Gly Thr Trp Leu Leu Asp Ser Glu Thr Met His Pro Leu Gly Met
            420                 425                 430

Ala Val Lys Ser Ser
        435
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 418 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Met Arg Leu Leu Thr Arg Val Leu Leu Val Ser Leu Leu Thr Leu Ser
 1               5                  10                  15

Leu Val Val Ser Gly Leu Ala Cys Gly Pro Gly Arg Gly Tyr Gly Arg
            20                  25                  30

Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile
        35                  40                  45

Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg Tyr Glu Gly
    50                  55                  60

Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn Tyr
65                  70                  75                  80

Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg
                85                  90                  95

Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ser Leu Ala Ile Ser
            100                 105                 110

Val Met Asn His Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp
        115                 120                 125

Asp Glu Asp Gly His His Phe Glu Glu Ser Leu His Tyr Glu Gly Arg
    130                 135                 140

Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Lys Ser Lys Tyr Gly Thr
145                 150                 155                 160

Leu Ser Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu
                165                 170                 175

Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val Ala
            180                 185                 190

Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Leu Val Ser Leu Gln
        195                 200                 205

Asp Gly Gly Gln Lys Ala Val Lys Asp Leu Asn Pro Gly Asp Lys Val
    210                 215                 220

Leu Ala Ala Asp Ser Ala Gly Asn Leu Val Phe Ser Asp Phe Ile Met
225                 230                 235                 240

Phe Thr Asp Arg Asp Ser Thr Thr Arg Arg Val Phe Tyr Val Ile Glu
                245                 250                 255

Thr Gln Glu Pro Val Glu Lys Ile Thr Leu Thr Ala Ala His Leu Leu
            260                 265                 270

Phe Val Leu Asp Asn Ser Thr Glu Asp Leu His Thr Met Thr Ala Ala
```

```
        275                 280                 285

Tyr Ala Ser Ser Val Arg Ala Gly Gln Lys Val Met Val Val Asp Asp
    290                 295                 300

Ser Gly Gln Leu Lys Ser Val Ile Val Gln Arg Ile Tyr Thr Glu Glu
305                 310                 315                 320

Gln Arg Gly Ser Phe Ala Pro Val Thr Ala His Gly Thr Ile Val Val
                325                 330                 335

Asp Arg Ile Leu Ala Ser Cys Tyr Ala Val Ile Glu Asp Gln Gly Leu
            340                 345                 350

Ala His Leu Ala Phe Ala Pro Ala Arg Leu Tyr Tyr Tyr Val Ser Ser
        355                 360                 365

Phe Leu Ser Pro Lys Thr Pro Ala Val Gly Pro Met Arg Leu Tyr Asn
    370                 375                 380

Arg Arg Gly Ser Thr Gly Thr Pro Gly Ser Cys His Gln Met Gly Thr
385                 390                 395                 400

Trp Leu Leu Asp Ser Asn Met Leu His Pro Leu Gly Met Ser Val Asn
                405                 410                 415

Ser Ser
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 475 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
Met Leu Leu Leu Ala Arg Cys Leu Leu Leu Val Leu Val Ser Ser Leu
 1               5                  10                  15

Leu Val Cys Ser Gly Leu Ala Cys Gly Pro Gly Arg Gly Phe Gly Lys
            20                  25                  30

Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys Gln Phe Ile
        35                  40                  45

Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Arg Tyr Glu Gly
    50                  55                  60

Lys Ile Ser Arg Asn Ser Glu Arg Phe Lys Glu Leu Thr Pro Asn Tyr
65                  70                  75                  80

Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Gly Ala Asp Arg
                85                  90                  95

Leu Met Thr Gln Arg Cys Lys Asp Lys Leu Asn Ala Leu Ala Ile Ser
```

```
                100                 105                 110

Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg Val Thr Glu Gly Trp
            115                 120                 125

Asp Glu Asp Gly His His Ser Glu Glu Ser Leu His Tyr Glu Gly Arg
        130                 135                 140

Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg Ser Lys Tyr Gly Met
145                 150                 155                 160

Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu
                165                 170                 175

Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn Ser Val Ala
            180                 185                 190

Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Ala Thr Val His Leu Glu
        195                 200                 205

Gln Gly Gly Thr Lys Leu Val Lys Asp Leu Ser Pro Gly Asp Arg Val
    210                 215                 220

Leu Ala Ala Asp Asp Gln Gly Arg Leu Leu Tyr Ser Asp Phe Leu Thr
225                 230                 235                 240

Phe Leu Asp Arg Asp Asp Gly Ala Lys Lys Val Phe Tyr Val Ile Glu
                245                 250                 255

Thr Arg Glu Pro Arg Glu Arg Leu Leu Leu Thr Ala Ala His Leu Leu
            260                 265                 270

Phe Val Ala Pro His Asn Asp Ser Ala Thr Gly Glu Pro Glu Ala Ser
        275                 280                 285

Ser Gly Ser Gly Pro Pro Ser Gly Gly Ala Leu Gly Pro Arg Ala Leu
    290                 295                 300

Phe Ala Ser Arg Val Arg Pro Gly Gln Arg Val Tyr Val Val Ala Glu
305                 310                 315                 320

Arg Asp Gly Asp Arg Arg Leu Leu Pro Ala Ala Val His Ser Val Thr
                325                 330                 335

Leu Ser Glu Glu Ala Ala Gly Ala Tyr Ala Pro Leu Thr Ala Gln Gly
            340                 345                 350

Thr Ile Leu Ile Asn Arg Val Leu Ala Ser Cys Tyr Ala Val Ile Glu
        355                 360                 365

Glu His Ser Trp Ala His Arg Ala Phe Ala Pro Phe Arg Leu Ala His
    370                 375                 380

Ala Leu Leu Ala Ala Leu Ala Pro Ala Arg Thr Asp Arg Gly Gly Asp
385                 390                 395                 400

Ser Gly Gly Gly Asp Arg Gly Gly Gly Gly Gly Arg Val Ala Leu Thr
                405                 410                 415

Ala Pro Gly Ala Ala Asp Ala Pro Gly Ala Gly Ala Thr Ala Gly Ile
            420                 425                 430
```

```
His Trp Tyr Ser Gln Leu Leu Tyr Gln Ile Gly Thr Trp Leu Leu Asp
        435                 440                 445

Ser Glu Ala Leu His Pro Leu Gly Met Ala Val Lys Ser Ser Xaa Ser
    450                 455                 460

Arg Gly Ala Gly Gly Gly Ala Arg Glu Gly Ala
465                 470                 475
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 411 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
Met Ser Pro Ala Arg Leu Arg Pro Arg Leu His Phe Cys Leu Val Leu
  1               5                  10                  15

Leu Leu Leu Leu Val Val Pro Ala Ala Trp Gly Cys Gly Pro Gly Arg
             20                  25                  30

Val Val Gly Ser Arg Arg Arg Pro Pro Arg Lys Leu Val Pro Leu Ala
         35                  40                  45

Tyr Lys Gln Phe Ser Pro Asn Val Pro Glu Lys Thr Leu Gly Ala Ser
     50                  55                  60

Gly Arg Tyr Glu Gly Lys Ile Ala Arg Ser Ser Glu Arg Phe Lys Glu
 65                  70                  75                  80

Leu Thr Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn
                 85                  90                  95

Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys Asp Arg Leu Asn
            100                 105                 110

Ser Leu Ala Ile Ser Val Met Asn Gln Trp Pro Gly Val Lys Leu Arg
        115                 120                 125

Val Thr Glu Gly Trp Asp Glu Asp Gly His His Ser Glu Glu Ser Leu
    130                 135                 140

His Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Arg
145                 150                 155                 160

Asn Lys Tyr Gly Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp
                165                 170                 175

Trp Val Tyr Tyr Glu Ser Lys Ala His Val His Cys Ser Val Lys Ser
            180                 185                 190

Glu His Ser Ala Ala Ala Lys Thr Gly Gly Cys Phe Pro Ala Gly Ala
        195                 200                 205

Gln Val Arg Leu Glu Ser Gly Ala Arg Val Ala Leu Ser Ala Val Arg
    210                 215                 220
```

Pro Gly Asp Arg Val Leu Ala Met Gly Glu Asp Gly Ser Pro Thr Phe
225 230 235 240

Ser Asp Val Leu Ile Phe Leu Asp Arg Glu Pro His Arg Leu Arg Ala
245 250 255

Phe Gln Val Ile Glu Thr Gln Asp Pro Pro Arg Arg Leu Ala Leu Thr
260 265 270

Pro Ala His Leu Leu Phe Thr Ala Asp Asn His Thr Glu Pro Ala Ala
275 280 285

Arg Phe Arg Ala Thr Phe Ala Ser His Val Gln Pro Gly Gln Tyr Val
290 295 300

Leu Val Ala Gly Val Pro Gly Leu Gln Pro Ala Arg Val Ala Ala Val
305 310 315 320

Ser Thr His Val Ala Leu Gly Ala Tyr Ala Pro Leu Thr Lys His Gly
325 330 335

Thr Leu Val Val Glu Asp Val Val Ala Ser Cys Phe Ala Ala Val Ala
340 345 350

Asp His His Leu Ala Gln Leu Ala Phe Trp Pro Leu Arg Leu Phe His
355 360 365

Ser Leu Ala Trp Gly Ser Trp Thr Pro Gly Glu Gly Val His Trp Tyr
370 375 380

Pro Gln Leu Leu Tyr Arg Leu Gly Arg Leu Leu Leu Glu Glu Gly Ser
385 390 395 400

Phe His Pro Leu Gly Met Ser Gly Ala Gly Ser
405 410

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 396 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Ala Leu Leu Thr Asn Leu Leu Pro Leu Cys Cys Leu Ala Leu Leu
1 5 10 15

Ala Leu Pro Ala Gln Ser Cys Gly Pro Gly Arg Gly Pro Val Gly Arg
20 25 30

Arg Arg Tyr Ala Arg Lys Gln Leu Val Pro Leu Leu Tyr Lys Gln Phe
35 40 45

Val Pro Gly Val Pro Glu Arg Thr Leu Gly Ala Ser Gly Pro Ala Glu
50 55 60

Gly Arg Val Ala Arg Gly Ser Glu Arg Phe Arg Asp Leu Val Pro Asn
65 70 75 80

Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Ser Gly Ala Asp
85 90 95

Arg Leu Met Thr Glu Arg Cys Lys Glu Arg Val Asn Ala Leu Ala Ile
100 105 110

Ala Val Met Asn Met Trp Pro Gly Val Arg Leu Arg Val Thr Glu Gly
115 120 125

Trp Asp Glu Asp Gly His His Ala Gln Asp Ser Leu His Tyr Glu Gly
130 135 140

Arg Ala Leu Asp Ile Thr Thr Ser Asp Arg Asp Arg Asn Lys Tyr Gly
145 150 155 160

Leu Leu Ala Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr
165 170 175

Glu Ser Arg Asn His Val His Val Ser Val Lys Ala Asp Asn Ser Leu
180 185 190

Ala Val Arg Ala Gly Gly Cys Phe Pro Gly Asn Ala Thr Val Arg Leu
195 200 205

Trp Ser Gly Glu Arg Lys Gly Leu Arg Glu Leu His Arg Gly Asp Trp
210 215 220

Val Leu Ala Ala Asp Ala Ser Gly Arg Val Val Pro Thr Pro Val Leu
225 230 235 240

Leu Phe Leu Asp Arg Asp Leu Gln Arg Arg Ala Ser Phe Val Ala Val
245 250 255

Glu Thr Glu Trp Pro Pro Arg Lys Leu Leu Leu Thr Pro Trp His Leu
260 265 270

Val Phe Ala Ala Arg Gly Pro Ala Pro Ala Pro Gly Asp Phe Ala Pro
275 280 285

Val Phe Ala Arg Arg Leu Arg Ala Gly Asp Ser Val Leu Ala Pro Gly
290 295 300

Gly Asp Ala Leu Arg Pro Ala Arg Val Ala Arg Val Ala Arg Glu Glu
305 310 315 320

Ala Val Gly Val Phe Ala Pro Leu Thr Ala His Gly Thr Leu Leu Val
325 330 335

Asn Asp Val Leu Ala Ser Cys Tyr Ala Val Leu Glu Ser His Gln Trp
340 345 350

Ala His Arg Ala Phe Ala Pro Leu Arg Leu Leu His Ala Leu Gly Ala
355 360 365

Leu Leu Pro Gly Gly Ala Val Gln Pro Thr Gly Met His Trp Tyr Ser
370 375 380

Arg Leu Leu Tyr Arg Leu Ala Glu Glu Leu Leu Gly
385 390 395

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 416 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Met Asp Val Arg Leu His Leu Lys Gln Phe Ala Leu Leu Cys Phe Ile
  1               5                  10                  15

Ser Leu Leu Leu Thr Pro Cys Gly Leu Ala Cys Gly Pro Gly Arg Gly
             20                  25                  30

Tyr Gly Lys Arg Arg His Pro Lys Lys Leu Thr Pro Leu Ala Tyr Lys
         35                  40                  45

Gln Phe Ile Pro Asn Val Ala Glu Lys Thr Leu Gly Ala Ser Gly Lys
     50                  55                  60

Tyr Glu Gly Lys Ile Thr Arg Asn Ser Glu Arg Phe Lys Glu Leu Ile
 65                  70                  75                  80

Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys Asp Glu Glu Asn Thr Asn
                 85                  90                  95

Ala Asp Arg Leu Met Thr Lys Arg Cys Lys Asp Lys Leu Asn Ser Leu
            100                 105                 110

Ala Ile Ser Val Met Asn His Trp Pro Gly Val Lys Leu Arg Val Thr
        115                 120                 125

Glu Gly Trp Asp Glu Asp Gly His His Leu Glu Glu Ser Leu His Tyr
    130                 135                 140

Glu Gly Arg Ala Val Asp Ile Thr Thr Ser Asp Arg Asp Lys Ser Lys
145                 150                 155                 160

Tyr Gly Met Leu Ser Arg Leu Ala Val Glu Ala Gly Phe Asp Trp Val
                165                 170                 175

Tyr Tyr Glu Ser Lys Ala His Ile His Cys Ser Val Lys Ala Glu Asn
            180                 185                 190

Ser Val Ala Ala Lys Ser Gly Gly Cys Phe Pro Gly Ser Gly Thr Val
        195                 200                 205

Thr Leu Gly Asp Gly Thr Arg Lys Pro Ile Lys Asp Leu Lys Val Gly
    210                 215                 220

Asp Arg Val Leu Ala Ala Asp Glu Lys Gly Asn Val Leu Ile Ser Asp
225                 230                 235                 240

Phe Ile Met Phe Ile Asp His Asp Pro Thr Thr Arg Arg Gln Phe Ile
                245                 250                 255

Val Ile Glu Thr Ser Glu Pro Phe Thr Lys Leu Thr Leu Thr Ala Ala
            260                 265                 270
```

His Leu Val Phe Val Gly Asn Ser Ser Ala Ala Ser Gly Ile Thr Ala
275 280 285

Thr Phe Ala Ser Asn Val Lys Pro Gly Asp Thr Val Leu Val Trp Glu
290 295 300

Asp Thr Cys Glu Ser Leu Lys Ser Val Thr Val Lys Arg Ile Tyr Thr
305 310 315 320

Glu Glu His Glu Gly Ser Phe Ala Pro Val Thr Ala His Gly Thr Ile
325 330 335

Ile Val Asp Gln Val Leu Ala Ser Cys Tyr Ala Val Ile Glu Asn His
340 345 350

Lys Trp Ala His Trp Ala Phe Ala Pro Val Arg Leu Cys His Lys Leu
355 360 365

Met Thr Trp Leu Phe Pro Ala Arg Glu Ser Asn Val Asn Phe Gln Glu
370 375 380

Asp Gly Ile His Trp Tyr Ser Asn Met Leu Phe His Ile Gly Ser Trp
385 390 395 400

Leu Leu Asp Arg Asp Ser Phe His Pro Leu Gly Ile Leu His Leu Ser
405 410 415

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1416 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: both
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1413

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

ATG GAT AAC CAC AGC TCA GTG CCT TGG GCC AGT GCC GCC AGT GTC ACC
48
Met Asp Asn His Ser Ser Val Pro Trp Ala Ser Ala Ala Ser Val Thr
1 5 10 15

TGT CTC TCC CTG GGA TGC CAA ATG CCA CAG TTC CAG TTC CAG TTC CAG
96
Cys Leu Ser Leu Gly Cys Gln Met Pro Gln Phe Gln Phe Gln Phe Gln
20 25 30

CTC CAA ATC CGC AGC GAG CTC CAT CTC CGC AAG CCC GCA AGA AGA ACG
144
Leu Gln Ile Arg Ser Glu Leu His Leu Arg Lys Pro Ala Arg Arg Thr
35 40 45

```
CAA ACG ATG CGC CAC ATT GCG CAT ACG CAG CGT TGC CTC AGC AGG CTG
192
Gln Thr Met Arg His Ile Ala His Thr Gln Arg Cys Leu Ser Arg Leu
     50                  55                  60

ACC TCT CTG GTG GCC CTG CTG CTG ATC GTC TTG CCG ATG GTC TTT AGC
240
Thr Ser Leu Val Ala Leu Leu Leu Ile Val Leu Pro Met Val Phe Ser
 65                  70                  75                  80

CCG GCT CAC AGC TGC GGT CCT GGC CGA GGA TTG GGT CGT CAT AGG GCG
288
Pro Ala His Ser Cys Gly Pro Gly Arg Gly Leu Gly Arg His Arg Ala
                 85                  90                  95

CGC AAC CTG TAT CCG CTG GTC CTC AAG CAG ACA ATT CCC AAT CTA TCC
336
Arg Asn Leu Tyr Pro Leu Val Leu Lys Gln Thr Ile Pro Asn Leu Ser
            100                 105                 110

GAG TAC ACG AAC AGC GCC TCC GGA CCT CTG GAG GGT GTG ATC CGT CGG
384
Glu Tyr Thr Asn Ser Ala Ser Gly Pro Leu Glu Gly Val Ile Arg Arg
        115                 120                 125

GAT TCG CCC AAA TTC AAG GAC CTC GTG CCC AAC TAC AAC AGG GAC ATC
432
Asp Ser Pro Lys Phe Lys Asp Leu Val Pro Asn Tyr Asn Arg Asp Ile
    130                 135                 140

CTT TTC CGT GAC GAG GAA GGC ACC GGA GCG GAT GGC TTG ATG AGC AAG
480
Leu Phe Arg Asp Glu Glu Gly Thr Gly Ala Asp Gly Leu Met Ser Lys
145                 150                 155                 160

CGC TGC AAG GAG AAG CTA AAC GTG CTG GCC TAC TCG GTG ATG AAC GAA
528
Arg Cys Lys Glu Lys Leu Asn Val Leu Ala Tyr Ser Val Met Asn Glu
                165                 170                 175

TGG CCC GGC ATC CGG CTG CTG GTC ACC GAG AGC TGG GAC GAG GAC TAC
576
Trp Pro Gly Ile Arg Leu Leu Val Thr Glu Ser Trp Asp Glu Asp Tyr
            180                 185                 190

CAT CAC GGC CAG GAG TCG CTC CAC TAC GAG GGC CGA GCG GTG ACC ATT
624
His His Gly Gln Glu Ser Leu His Tyr Glu Gly Arg Ala Val Thr Ile
        195                 200                 205

GCC ACC TCC GAT CGC GAC CAG TCC AAA TAC GGC ATG CTC GCT CGC CTG
672
Ala Thr Ser Asp Arg Asp Gln Ser Lys Tyr Gly Met Leu Ala Arg Leu
    210                 215                 220

GCC GTC GAG GCT GGA TTC GAT TGG GTC TCC TAC GTC AGC AGG CGC CAC
720
Ala Val Glu Ala Gly Phe Asp Trp Val Ser Tyr Val Ser Arg Arg His
225                 230                 235                 240

ATC TAC TGC TCC GTC AAG TCA GAT TCG TCG ATC AGT TCC CAC GTG CAC
768
```

```
Ile Tyr Cys Ser Val Lys Ser Asp Ser Ser Ile Ser Ser His Val His
                245                     250                     255

GGC TGC TTC ACG CCG GAG AGC ACA GCG CTG CTG GAG AGT GGA GTC CGG
816
Gly Cys Phe Thr Pro Glu Ser Thr Ala Leu Leu Glu Ser Gly Val Arg
            260                     265                     270

AAG CCG CTC GGC GAG CTC TCT ATC GGA GAT CGT GTT TTG AGC ATG ACC
864
Lys Pro Leu Gly Glu Leu Ser Ile Gly Asp Arg Val Leu Ser Met Thr
        275                     280                     285

GCC AAC GGA CAG GCC GTC TAC AGC GAA GTG ATC CTC TTC ATG GAC CGC
912
Ala Asn Gly Gln Ala Val Tyr Ser Glu Val Ile Leu Phe Met Asp Arg
    290                     295                     300

AAC CTC GAG CAG ATG CAA AAC TTT GTG CAG CTG CAC ACG GAC GGT GGA
960
Asn Leu Glu Gln Met Gln Asn Phe Val Gln Leu His Thr Asp Gly Gly
305                     310                     315                     320

GCA GTG CTC ACG GTG ACG CCG GCT CAC CTG GTT AGC GTT TGG CAG CCG
1008
Ala Val Leu Thr Val Thr Pro Ala His Leu Val Ser Val Trp Gln Pro
                325                     330                     335

GAG AGC CAG AAG CTC ACG TTT GTG TTT GCG CAT CGC ATC GAG GAG AAG
1056
Glu Ser Gln Lys Leu Thr Phe Val Phe Ala His Arg Ile Glu Glu Lys
            340                     345                     350

AAC CAG GTG CTC GTA CGG GAT GTG GAG ACG GGC GAG CTG AGG CCC CAG
1104
Asn Gln Val Leu Val Arg Asp Val Glu Thr Gly Glu Leu Arg Pro Gln
        355                     360                     365

CGA GTG GTC AAG TTG GGC AGT GTG CGC AGT AAG GGC GTG GTC GCG CCG
1152
Arg Val Val Lys Leu Gly Ser Val Arg Ser Lys Gly Val Val Ala Pro
    370                     375                     380

CTG ACC CGC GAG GGC ACC ATT GTG GTC AAC TCG GTG GCC GCC AGT TGC
1200
Leu Thr Arg Glu Gly Thr Ile Val Val Asn Ser Val Ala Ala Ser Cys
385                     390                     395                     400

TAT GCG GTG ATC AAC AGT CAG TCG CTG GCC CAC TGG GGA CTG GCT CCC
1248
Tyr Ala Val Ile Asn Ser Gln Ser Leu Ala His Trp Gly Leu Ala Pro
                405                     410                     415

ATG CGC CTG CTG TCC ACG CTG GAG GCG TGG CTG CCC GCC AAG GAG CAG
1296
Met Arg Leu Leu Ser Thr Leu Glu Ala Trp Leu Pro Ala Lys Glu Gln
            420                     425                     430

TTG CAC AGT TCG CCG AAG GTG GTG AGC TCG GCG CAG CAG CAG AAT GGC
1344
Leu His Ser Ser Pro Lys Val Val Ser Ser Ala Gln Gln Gln Asn Gly
        435                     440                     445
```

```
ATC CAT TGG TAT GCC AAT GCG CTC TAC AAG GTC AAG GAC TAC GTG CTG  1392
Ile His Trp Tyr Ala Asn Ala Leu Tyr Lys Val Lys Asp Tyr Val Leu
    450                 455                 460

CCG CAG AGC TGG CGC CAC GAT TGA  1416
Pro Gln Ser Trp Arg His Asp
465                 470
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 471 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Met Asp Asn His Ser Ser Val Pro Trp Ala Ser Ala Ala Ser Val Thr
  1               5                  10                  15

Cys Leu Ser Leu Gly Cys Gln Met Pro Gln Phe Gln Phe Gln Phe Gln
             20                  25                  30

Leu Gln Ile Arg Ser Glu Leu His Leu Arg Lys Pro Ala Arg Arg Thr
         35                  40                  45

Gln Thr Met Arg His Ile Ala His Thr Gln Arg Cys Leu Ser Arg Leu
     50                  55                  60

Thr Ser Leu Val Ala Leu Leu Leu Ile Val Leu Pro Met Val Phe Ser
 65                  70                  75                  80

Pro Ala His Ser Cys Gly Pro Gly Arg Gly Leu Gly Arg His Arg Ala
                 85                  90                  95

Arg Asn Leu Tyr Pro Leu Val Leu Lys Gln Thr Ile Pro Asn Leu Ser
            100                 105                 110

Glu Tyr Thr Asn Ser Ala Ser Gly Pro Leu Glu Gly Val Ile Arg Arg
        115                 120                 125

Asp Ser Pro Lys Phe Lys Asp Leu Val Pro Asn Tyr Asn Arg Asp Ile
    130                 135                 140

Leu Phe Arg Asp Glu Glu Gly Thr Gly Ala Asp Gly Leu Met Ser Lys
145                 150                 155                 160

Arg Cys Lys Glu Lys Leu Asn Val Leu Ala Tyr Ser Val Met Asn Glu
                165                 170                 175

Trp Pro Gly Ile Arg Leu Leu Val Thr Glu Ser Trp Asp Glu Asp Tyr
            180                 185                 190

His His Gly Gln Glu Ser Leu His Tyr Glu Gly Arg Ala Val Thr Ile
        195                 200                 205
```

```
Ala Thr Ser Asp Arg Asp Gln Ser Lys Tyr Gly Met Leu Ala Arg Leu
    210                 215                 220

Ala Val Glu Ala Gly Phe Asp Trp Val Ser Tyr Val Ser Arg Arg His
225                 230                 235                     240

Ile Tyr Cys Ser Val Lys Ser Asp Ser Ser Ile Ser Ser His Val His
                245                 250                 255

Gly Cys Phe Thr Pro Glu Ser Thr Ala Leu Leu Glu Ser Gly Val Arg
            260                 265                 270

Lys Pro Leu Gly Glu Leu Ser Ile Gly Asp Arg Val Leu Ser Met Thr
        275                 280                 285

Ala Asn Gly Gln Ala Val Tyr Ser Glu Val Ile Leu Phe Met Asp Arg
    290                 295                 300

Asn Leu Glu Gln Met Gln Asn Phe Val Gln Leu His Thr Asp Gly Gly
305                 310                 315                     320

Ala Val Leu Thr Val Thr Pro Ala His Leu Val Ser Val Trp Gln Pro
                325                 330                 335

Glu Ser Gln Lys Leu Thr Phe Val Phe Ala His Arg Ile Glu Glu Lys
            340                 345                 350

Asn Gln Val Leu Val Arg Asp Val Glu Thr Gly Glu Leu Arg Pro Gln
        355                 360                 365

Arg Val Val Lys Leu Gly Ser Val Arg Ser Lys Gly Val Val Ala Pro
    370                 375                 380

Leu Thr Arg Glu Gly Thr Ile Val Val Asn Ser Val Ala Ala Ser Cys
385                 390                 395                     400

Tyr Ala Val Ile Asn Ser Gln Ser Leu Ala His Trp Gly Leu Ala Pro
                405                 410                 415

Met Arg Leu Leu Ser Thr Leu Glu Ala Trp Leu Pro Ala Lys Glu Gln
            420                 425                 430

Leu His Ser Ser Pro Lys Val Val Ser Ser Ala Gln Gln Gln Asn Gly
        435                 440                 445

Ile His Trp Tyr Ala Asn Ala Leu Tyr Lys Val Lys Asp Tyr Val Leu
    450                 455                 460

Pro Gln Ser Trp Arg His Asp
465                 470
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
Cys Gly Pro Gly Arg Gly Xaa Gly Xaa Arg Arg His Pro Lys Lys Leu
1               5                   10                  15

Thr Pro Leu Ala Tyr Lys Gln Phe Ile Pro Asn Val Ala Glu Lys Thr
            20                  25                  30

Leu Gly Ala Ser Gly Arg Tyr Glu Gly Lys Ile Xaa Arg Asn Ser Glu
        35                  40                  45

Arg Phe Lys Glu Leu Thr Pro Asn Tyr Asn Pro Asp Ile Ile Phe Lys
    50                  55                  60

Asp Glu Glu Asn Thr Gly Ala Asp Arg Leu Met Thr Gln Arg Cys Lys
65                  70                  75                  80

Asp Lys Leu Asn Xaa Leu Ala Ile Ser Val Met Asn Xaa Trp Pro Gly
                85                  90                  95

Val Xaa Leu Arg Val Thr Glu Gly Trp Asp Glu Asp Gly His His Xaa
            100                 105                 110

Glu Glu Ser Leu His Tyr Glu Gly Arg Ala Val Asp Ile Thr Thr Ser
        115                 120                 125

Asp Arg Asp Xaa Ser Lys Tyr Gly Xaa Leu Xaa Arg Leu Ala Val Glu
    130                 135                 140

Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Lys Ala His Ile His Cys
145                 150                 155                 160

Ser Val Lys Ala Glu Asn Ser Val Ala Ala Lys Ser Gly Gly Cys Phe
                165                 170                 175

Pro Gly Ser Ala Xaa Val Xaa Leu Xaa Xaa Gly Gly Xaa Lys Xaa Val
            180                 185                 190

Lys Asp Leu Xaa Pro Gly Asp Xaa Val Leu Ala Ala Asp Xaa Xaa Gly
        195                 200                 205

Xaa Leu Xaa Xaa Ser Asp Phe Xaa Xaa Phe Xaa Asp Arg
    210                 215                 220
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 167 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Cys Gly Pro Gly Arg Gly Xaa Xaa Xaa Arg Arg Xaa Xaa Xaa Pro Lys
1               5                   10                  15

Xaa Leu Xaa Pro Leu Xaa Tyr Lys Gln Phe Xaa Pro Xaa Xaa Xaa Glu
            20                  25                  30

Xaa Thr Leu Gly Ala Ser Gly Xaa Xaa Glu Gly Xaa Xaa Xaa Arg Xaa
        35                  40                  45

Ser Glu Arg Phe Xaa Xaa Leu Thr Pro Asn Tyr Asn Pro Asp Ile Ile
    50                  55                  60

Phe Lys Asp Glu Glu Asn Xaa Gly Ala Asp Arg Leu Met Thr Xaa Arg
65                  70                  75                  80

Cys Lys Xaa Xaa Xaa Asn Xaa Leu Ala Ile Ser Val Met Asn Xaa Trp
                85                  90                  95

Pro Gly Val Xaa Leu Arg Val Thr Glu Gly Xaa Asp Glu Asp Gly His
            100                 105                 110

His Xaa Xaa Xaa Ser Leu His Tyr Glu Gly Arg Ala Xaa Asp Ile Thr
        115                 120                 125

Thr Ser Asp Arg Asp Xaa Xaa Lys Tyr Gly Xaa Leu Xaa Arg Leu Ala
    130                 135                 140

Val Glu Ala Gly Phe Asp Trp Val Tyr Tyr Glu Ser Xaa Xaa His Xaa
145                 150                 155                 160

His Xaa Ser Val Lys Xaa Xaa
                165
```